IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 19, 2021, accepted June 28, 2021, date of publication July 2, 2021, date of current version August 2, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3094196

A Survey on Centrality Metrics and Their Network Resilience Analysis

ZELIN WAN[1], YASH MAHAJAN[1], BEOM WOO KANG[2], TERRENCE J. MOORE[3], (Member, IEEE), AND JIN-HEE CHO[1], (Senior Member, IEEE)

[1]Department of Computer Science, Virginia Tech, Blacksburg, VA 24061, USA
[2]Department of Electronic Engineering, Hanyang University, Seoul 04763, South Korea
[3]U.S. Army Research Laboratory, Adelphi, MD 20783, USA

Corresponding author: Zelin Wan (zelin@vt.edu)

ABSTRACT Centrality metrics have been studied in the network science research. They have been used in various networks, such as communication, social, biological, geographic, or contact networks under different disciplines. In particular, centrality metrics have been used in order to study and analyze targeted attack behaviors and investigated their effect on network resilience. Although a rich volume of centrality metrics has been developed from 1940s, only some centrality metrics (e.g., degree, betweenness, or cluster coefficient) have been commonly in use. This paper aims to introduce various existing centrality metrics and discusses their applicabilities in various networks. In addition, we conducted extensive simulation study in order to demonstrate and analyze the network resilience of targeted attacks using the surveyed centrality metrics under four real network topologies. We also discussed algorithmic complexity of centrality metrics surveyed in this work. Through the extensive experiments and discussions of the surveyed centrality metrics, we encourage their use in solving various computing and engineering problems in networks.

INDEX TERMS Centrality, networks, influence, importance, attacks, network resilience, network science.

I. INTRODUCTION

A. MOTIVATION

Identifying central nodes in a network is critical to the design of a network that is resilient against faults or attacks. However, identifying which nodes are vital in a network is a non-trivial task. Centrality has been studied in a fashion since the beginnings of graph theory, where the degree is central to Euler's proof of the impossibility of satisfying the Seven Bridges of Königsberg problem in the 1700s [1]. A significant spike of interest in centrality occurred in the late 1940s and early 1950s [2]–[7] in the social sciences. A more interdisciplinary approach emerged in the late 1990s and the early 2000s in the nomenclature of *Network Science* [8]. In the resilience context, there is extensive literature studying the effect of targeted attacks, or attacks on nodes that have high centrality [9]–[13]. A typical scenario includes an intelligent attacker that selects a target node or nodes to disrupt or compromise the network. Since the 2000s, centrality metrics have grown in significance in communication networks as network resilience and cybersecurity concerns have become more prominent. The most common centrality metrics are degree [9], [14], betweenness [15], [16], eigenvector [17], closeness [18], pagerank [19], and so forth. However, given the rich volume of existing centrality metrics studied in other scientific fields for decades, their merits and relevant usages have been insufficiently appreciated and leveraged in various communication and network domains. In this survey, we aim to present this rich volume of centrality metrics and how they can be used in various network and communication research. In addition, we demonstrate the performance of each centrality metric in terms of the effect of targeted attack based on the metrics on the resilience of several real-world networks. We hope this work can open a door for researchers in engineering fields to fully leverage the existing centrality metrics and their relevancy for network system design and attack modeling.

B. COMPARISON OF OUR SURVEY PAPER AND EXISTING CENTRALITY METRICS SURVEY PAPERS

The study of centrality metrics has a history stretching back nearly a century. In the past two decades, many

The associate editor coordinating the review of this manuscript and approving it for publication was Chao-Yang Chen.

comprehensive surveys have attempted to catalog and understand these metrics. In 2002, Dhyani *et al.* [20] conducted a survey on metrics used in Web information networks to measure graph properties, page importance, page similarity, search, retrieval, the characteristics of usage, and information-theoretic properties. Other fairly recent efforts surveyed centrality applicable in multiple domains. For example, Guille *et al.* [21] surveyed a small set of centrality metrics and tested their impact on information diffusion in terms of topic propagation originating at those central nodes. This work is limited to the application of information diffusion with over a dozen well-known centrality metrics. Lü *et al.* [22] conducted a more comprehensive survey on centrality metrics and demonstrated their performance in various network types. The authors considered biological networks, financial networks, social networks, and software networks. The authors also studied different types of networks, such as directed, undirected, bipartite, and weighted networks. Their performance analysis of applications included the investigation of centrality on information diffusion, identification of scientific influence, detection of financial risks, prediction of essential proteins, and so forth. However, network resilience has not been considered in their analysis, which is a central theme of our paper. More recently, Das *et al.* [23] surveyed over a dozen centrality metrics applicable in the social networks context, providing a nice discussion of the linkages between the centralities they discussed. Ashtiani *et al.* [24] conducted a comprehensive survey on centrality metrics to investigate protein-protein interaction networks. They examined node centrality in yeast protein-protein interaction networks (PPINs) for the detection or prediction of influential proteins. However, this work is also limited to the application of centrality metrics in biological networks. Lalou *et al.* [25] reviewed the recent studies that solved the problem of identifying critical nodes in a network. This work mainly focused on surveying theoretical complexity, exact solving algorithms (not centrality metrics), approximation schemes, and heuristic approaches. In particular, the authors proved new complexity results of algorithms that were improved based on the relationships between variants.

Unlike these prior surveys [20]–[25], our survey primarily focuses on the investigation of node centrality, graph centrality, and group-selection centrality in the context of the impact of centrality on network resilience under targeted attacks.

C. KEY CONTRIBUTIONS & SCOPE

Unlike the other state-of-the-art survey papers above, this survey paper makes the following **key contributions**:

- We discussed multidisciplinary concepts of centrality and its historical evolution in the research literature. This provides insights on how centrality metrics have been applied in various kinds of networks, in particular their applicability in communication and social networks of interest to many engineers.
- We conducted an extensive survey on three types of centrality metrics, consisting of point centrality metrics, graph centrality metrics, and group selection metrics, covering 60 centrality metrics in total. We also described how each metric is computed and its computational complexity. This may inform other researchers into what metric will be more relevant for a particular network or system design of interest.
- Unlike other conventional survey papers, we conducted extensive simulation experiments to demonstrate the performance of the surveyed centrality metrics in terms of network resilience based on the size of the giant component where each centrality metric is used to pick targets to model either non-infectious or infectious attacks. In addition, we demonstrated the simulation running time of each centrality metric and analyzed their algorithmic efficiency. The extensive discussions of the experimental results from this study will provide a clear and in-depth understanding of how one metric is more relevant than others based on a comparative performance analysis under four different real network topologies.
- Based on the extensive survey and experimental performance comparison of the centrality metrics, we shared what we have learned, providing both insights, limitations as well as promising future research directions.

D. PAPER STRUCTURE

The rest of this paper is organized as follows:

- Section II discusses the multidisciplinary concepts of centrality that have been studied in various domains studying networks. We also discuss how centrality metrics have been utilized in different disciplines as well as the evolution of research on centrality over time.
- Sections III–V provide an extensive survey on centrality metrics, categorized into three classes: point centrality (Section III) measuring the role of the node in the network, graph centrality (Section IV) measuring a characteristic of the graph as a whole, and group selection metrics (Section V) measuring the distributed spread of central nodes. We discuss the interpretation of each metric and describe its computation in the network. In addition, we discuss the algorithmic complexity of each metric, which can provide useful insights when researchers need to select a relevant centrality metric depending on the characteristics of the system of interest.
- Section VI provides an overview of how centrality metrics have evolved from prior to the 1970s up to the 2010s.
- Section VII demonstrates the performance of the surveyed centrality metrics in terms of their effect as targeted attacks on network resilience based on the size of the giant component. We implemented 56 metrics (i.e., 38 point centrality metrics, 13 graph centrality metrics, and 5 group selection metrics) and investigated their effect on network resilience. We used 4 different real

communication network topologies datasets, 2 directed networks, and 2 undirected networks, to investigate the effect of the centrality metrics surveyed in this work. In addition, we considered two types of targeted attacks, non-infectious attacks (wherein adjacent nodes of targeted nodes are not impacted) and infectious attacks (or 'epidemic attacks,' wherein adjacent nodes may be impacted and infected by targeted nodes).
- Section VIII discusses how existing centrality metrics have been applied to solve various problems in different types of networks.
- Section IX discusses the limitations of these state-of-the-art centrality metrics as well as provides insights and lessons learned from this extensive survey. And then, we finally conclude this paper with extensive discussions of promising future research directions.

II. MULTIDIMENSIONAL CONCEPTS OF CENTRALITY AND ITS APPLICATIONS IN DIVERSE DOMAINS

The multidisciplinary development of network centrality concepts has generated multifaceted interpretations of the subject. In this section, we discuss the different concepts of centrality that have been studied in different disciplines and specifically described their study in the network science context. We finish the section with common notations that are used in the succeeding sections describing centrality metrics.

A. MULTIDIMENSIONAL CONCEPTS OF CENTRALITY

A fundamental motivation for the study of centrality is the belief that one's position in the network impacts their access to information [4], [26], status [5], power [27], prestige [28], and influence [29]. We categorize these concepts into three classes as follows: (1) *communication activity* based on individual characteristics; (2) *influence* based on both individual and network characteristics; and (3) *communication control* based primarily on network characteristics. Individual characteristics refer to the way an individual node (i.e., user) interacts with other nodes, such as the frequency of interactions (e.g., posting or sending information in online social networks, namely OSNs, or sending signals or packets in communication networks), the amount of information sharing with others, and the quality of the signals (e.g., posted comments). Network characteristics typically indicate which nodes are connected with which nodes. It is these characteristics that can be captured by centrality.

1) COMMUNICATION ACTIVITY

This aspect of centrality covers the amount and type of activity an individual node participates in as part of its communications with other nodes. The activity of a node relative to other nodes can ultimately affect its power or influence. Klein *et al.* [30] demonstrated a connection between the communication activity and the influence of a user in an OSN. In OSNs, influential users tend to more easily spread information they choose to communicate. However, such well-connected users are less likely to disseminate information received from their extensive network. Hence, this characteristic in terms of frequency or type of interactions of information sharing is a critical factor related to centrality [31].

2) INFLUENCE

Influence has been used to represent the capability to affect other nodes. A number of terms are used to characterize and study the 'influence' of a node as follows:

- *Power:* Friedkin [29] examined the relationships between network centrality and the mutual influence of members in a group. An individual member's centrality affects other members' opinions and informs a dynamic process of updating their opinions.
- *Status:* Katz [5] proposed the idea that a member's centrality within a network depends upon not only the number of adjacent neighbors but also the *status* of each neighbor, i.e., the highest-status member who obtains the majority of votes from other nodes in a network becomes the most influential. Katz introduced an advanced metric to calculate the status of each member in a network based on the total number of votes, implying the edges in a directed graph, toward each member via a single step up to multiple steps that entail attenuation in a connection of a series [32].
- *Prestige:* In a social network, a person's prestige, which is often measured by 'in-degree', indicates the extent to which other people reach out to the person [33]. A person with high prestige refers to an influencer as the object of communication, rather than the source of communication [34]. Bonacich [27] and Katz [5] defined a vertex's prestige in a network based on its neighbors. For example, eigenvector centrality is used to derive the prestige of each vertex [28].
- *Resources:* How much resource one can obtain from their network has been discussed within the context of an exchange network [35]. In an exchange network, consisting of a set of members exchanging opportunities, each member needs to decide whether to connect with others to increase their opportunities or resources even when unaware of members outside of its own set of exchanging opportunities [36]. This feature facilitates the analysis of the power distribution as related to the position in the network [35], [36]. In exchange networks, a node's power is not necessarily aligned with the number of connections [35]. Centrality metrics that are more relevant to quick-spreading or mitigating influence (e.g., information diffusion or disease transmission) are more reliant on the number of direct or indirect connections with other nodes. Bonacich [27] reflected this belief in his eigenvector-type centrality where a node's power is measured based on the power of its neighbors. Laumann and Pappi [37] discussed a *community elite*, a set of necessary members in exchange networks in

which their position and other attributes determine the structure of influence.

- *Bridging:* Saito *et al.* [38] introduced the concept of *super-mediators* as the set of nodes that transfer information between nodes. The capability of a certain node to receive information from numerous nodes and propagate this information to others indicates their influence [4], [26]. Betweenness [13], [39] is an example representing a bridging role in a network where the node with high betweenness can connect other nodes as a key mediator. This concept of a broker in sociology is commonly described as a node with high betweenness that can play a key role in bridging two separate groups, which is often explained as a key factor in social capital [13].

3) COMMUNICATION CONTROL

A node's communication control describes how the node can control communications with others, which can naturally affect the node's centrality. The common two factors affecting this communication control are:

- *Commnicability:* With respect to group performance and individual behavioral patterns, Leavitt [4] stressed the importance of a network topology because it determines information accessibility that can affect successful task executions.
- *Network size:* A network can be viewed as a resource as each individual gathers information via connections within networks [40]. A node's network size is a typical measure of the node's centrality in terms of the resources available to it, including both the quality and the quantity of information in its network [41].

B. CENTRALITY METRICS RESEARCH IN VARIOUS DISCIPLINES

The study of centrality metrics has a long history from 1940s and has been conducted in various disciplines: mathematics [42], chemistry [43]–[45], anthropology [2], [6], [46], [47], physics [8], [13], geography [48], [49], economics [50], [51], psychology [52]–[54], sociology [27], [55]–[60], biology [61]–[63], management [64]–[66], computer science [67]–[70], political science [71]–[73], and psychiatry [72]–[75].

Fig. 1 summarizes the evolution of centrality across diverse disciplines along with the emergence of the *Network Science* discipline. The origin of developing centrality metrics is linked with the birth of graph theory [42]. Although many fields have used centrality metrics for a variety of purposes, high visibility of the usefulness of these metrics has been much increased as the Network Science field has officially formed in 2000s. In particular, in 2006, US National Research Council (NRC) defined *Network Science* as an academic field [8]. In 2009, The Department of Defense (DoD) initiated a research effort on Network Science for developing battlefield platforms with advanced technology reflecting the theme of *network-centric warfare*. The US Army Research Laboratory (US-ARL) initiated a collaborative research program, the *Network Science Collaborative Technology Alliance* (NS CTA), in order to encourage the development of advanced network science-based technologies to support ground soldiers in network-centric warfare [76], which has further triggered the advancement and maturity of network science research.

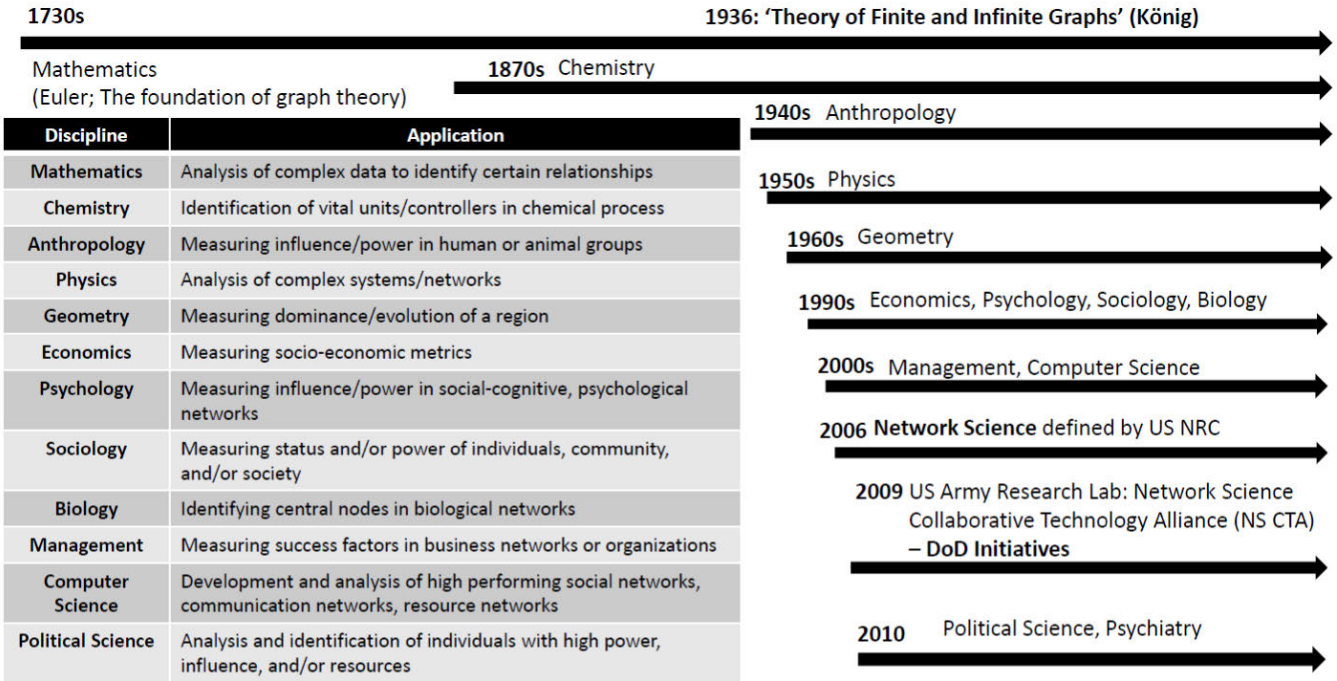
C. MEASURES OF NETWORK RESILIENCE IN NETWORK SCIENCE

Colbourn [77] defined the concept of 'network resilience' as ''the expected number of node pairs which can communicate'' in the presence of failed nodes. That is, the measure of network resilience is the probability of an operating path existing between two nodes. Najjar and Gaudiot [78] defined network resilience as the ''maximum number of node failures that can be sustained while the network remains connected'' when the failure rate is given. The concept of network resilience in the network science domain is very well aligned with these two definitions. Network science researchers often interchangeably use network resilience with network robustness. Most network science studies commonly use the size of the giant component, which indicates the size of a largest network component in a given network after attacks are applied. Hence, the measure of network resilience using the size of the giant component estimates the extent of 'fault-tolerance' based on the amount of topological connectivities in the given network [79].

Network resilience has been measured differently depending on different contexts. In particular, Rueda *et al.* [80] surveyed various metrics to measure structural robustness, which can indicate the ability to provide continuous operation in the presence of failures. Santos *et al.* [81] considered connectivity-based network resilience for software-defined networks aiming to solve a controller deployment problem. The network resilience or robustness metrics suggested by these two works [80], [81] are well aligned in that a resilient (or robust) network should be able to provide continuous network connectivity. Since the scope of this paper is to provide general trends of network resilience for various centrality metrics, we chose the conventional metric of network resilience, which is *the size of the giant component* for the network resilience analysis in Section VII. We also analyze the algorithmic complexity of the centrality metrics we surveyed in Section VII.

D. NOTATIONS

Before introducing various centrality metrics, we first present some common notations. We represent a network using a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is a set of n vertices representing the nodes in the network and $\mathcal{E}$ is the set of m edges representing connections (links or relationships) between pairs of nodes. All the graphs considered in this work are simple, meaning there are no pairs of nodes with multiple edges between them and there are no nodes that have self-loops



Discipline	Application
Mathematics	Analysis of complex data to identify certain relationships
Chemistry	Identification of vital units/controllers in chemical process
Anthropology	Measuring influence/power in human or animal groups
Physics	Analysis of complex systems/networks
Geometry	Measuring dominance/evolution of a region
Economics	Measuring socio-economic metrics
Psychology	Measuring influence/power in social-cognitive, psychological networks
Sociology	Measuring status and/or power of individuals, community, and/or society
Biology	Identifying central nodes in biological networks
Management	Measuring success factors in business networks or organizations
Computer Science	Development and analysis of high performing social networks, communication networks, resource networks
Political Science	Analysis and identification of individuals with high power, influence, and/or resources
Psychiatry	Investigating the relationships between disorder symptoms and disease in psychiatry networks

FIGURE 1. Evolution of network centrality metrics across multiple disciplines.

(no nodes with an edge to themselves). This graph can be represented in a matrix form using an adjacency matrix $\mathbf{A}$, where $A_{uv} \neq 0$ means there exists an edge between u and v (and the suffix uv of the matrix indicates the row entry corresponds to the node u and the column entry corresponds to the node v). For a simple, undirected graph, this value is 1 and the adjacency matrix is symmetric, i.e., $A_{uv} = A_{vu}$. For a weighted graph, the value of the adjacency entries corresponds to the weight. For a directed graph, the A_{uv} corresponds to the directed edge from u to v, and the matrix may not be symmetric. The Laplacian of a simple, undirected (and unweighted) graph is $\mathbf{L} = \mathbf{D} - \mathbf{A}$, where here $\mathbf{D}$ is a diagonal matrix with the ith diagonal entry equal to the sum of the ith row of $\mathbf{A}$. Many approaches rely on the principal eigenvector of the adjacency matrix or another matrix, so we denote $\mathbf{e}_1(\mathbf{M})$ to be this eigenvector of the matrix $\mathbf{M}$. The trace of a matrix $\mathbf{M}$, denoted $\mathrm{Tr}(\mathbf{M})$, is the sum of the diagonal elements of $\mathbf{M}$. We also use $\mathbf{I}$ and $\mathbf{1}$ to represent the identity matrix and a vector of ones, respectively.

Two key notions in many of the metrics are the distance between two vertices and the neighbors of a vertex. A path in the graph is any sequence of vertices such that an edge exists between successive vertices in the sequence. The distance between two vertices v and u, which we denote as $d(u, v)$, is the number of edges on the shortest path between the two vertices. The set of all nodes that are a distance 1 from a vertex v, denoted $N(v)$, is its neighbor set (or its adjacent nodes). The size of this set is denoted $d(v) = |N(v)|$ and is the degree of v. This concept can be extended for arbitrary distances, e.g., $N_h(v) = \{u : d(u, v) \leq h\}$ and $d_h(v) = |N_h(v)|$. For directed graphs, we distinguish between neighbors based on the direction of the edge; neighbors with directed edges from v are in the set $N^{\text{out}}(v)$ and neighbors with directed edges to v are in the set $N^{\text{in}}(v)$. We denote the average degree in the graph as $\langle k \rangle$. Other notations are defined in the context of the centrality for which they are used.

Now we discuss a variety of centrality metrics that have been proposed in various disciplines. We categorize the types of centrality metrics into three classes: point centrality metrics, graph centrality metrics, and group selection centrality metrics. Sections III-V below address these three classes of centrality metrics.

III. POINT CENTRALITY METRICS

We classify point centrality metrics in terms of three classes: local centrality metrics, iterative centrality metrics, and global centrality metrics.

A. MEASURES OF LOCAL CENTRALITY

Local centrality metrics measure the centrality of a node based on its local neighborhood topology. These metrics are variations of the degree of a node, sometimes in combination with the degree of other nodes in its local neighborhood.

1) DEGREE CENTRALITY

The simplest and most well-known centrality metric is the node degree or the number of links or edges incident to the node. The degree of vertex v is defined, mathematically, by[1]:

$$\begin{aligned} C_{\text{deg}}(v) &= \#\text{ of edges incident to } v \\ &= \sum_{u\in\mathcal{V}} A_{vu} = \sum_{u\in\mathcal{V}} A_{uv}. \end{aligned} \tag{1}$$

In the social network context, the degree indicates the number of relations of the actor [82], [83]. These relationships may be friendships [84], [85] or interactions [86] or other activities. Hanneman and Riddle [87] describe the degree as measuring the opportunities and alternatives for the actor. In social, communication, and computer networks, degree represents a measure of the number of channels for information or product exchange (i.e., sending and receiving data) [13], [88]. This information exchange can be messages in a radio network [89] or messages passed over the internet [90]. An example of a product exchange can be current in a power grid [91]. A standardization or normalization of degree is given by $C_{\text{deg}}(v)/(n-1)$. This form is useful for comparison across networks [82]. Nodes with a high degree are called *hubs*. In a directed network, the in-degree and the out-degree of the node may be unequal, so the adjacency matrix is not symmetric. For in-degree, $C_{\text{in-deg}}(v) =$ # edges directed toward $v = \sum_{u\in\mathcal{V}} A_{uv}$, with the out-degree defined analogously based on non-zero entries in the row of the adjacency matrix corresponding to node v. A node with significantly higher in-degree than out-degree or higher in-degree on average compared with other nodes is considered to have prestige [83]. A popular example exists in citation networks where the directed edges correspond to one document citing another. Documents with many citations have high in-degree. Other modified examples corresponding to in-degree in citations include the number of citations of a given author [92] and journal impact factor [93].

2) SEMI-LOCAL CENTRALITY

While a hub node has immediate access to a large number of neighbors, the hub may exist on the periphery or far from the center of the network where most of those neighbors have little to no access to the rest of the network. Hence, the degree alone may not be the ideal measure of influence or the capability of spreading (information or disease) with efficacy. Seeking a middle ground between hub nodes and nodes that have high betweenness (see Eq. (26)), Chen *et al.* [94] developed *semi-local centrality*, sometimes called *local centrality*, as a low-complexity approach that takes into account neighbor degrees of the node. The semi-local centrality of a node v is defined as:

$$C_{\text{semi-local}}(v) = \sum_{u\in N(v)} Q(u), \tag{2}$$

where $Q(u) = \sum_{w\in N(u)} d_2(w)$. This metric compares favorably to ranks generated from a susceptible-infected-removed (SIR) process [13].

3) HYBRID DEGREE CENTRALITY

In the context of spreading processes, whether it be for information sharing or disease transmission, the spreading probability p can determine the difference between the influence of the local and near-local neighborhood topology. A small p would intuitively favor a measure like degree centrality, while a larger p would favor a more global measure. Ma and Ma [95] incorporated the influence of the scale of p into centrality by adapting degree centrality and semi-local centrality [94] to create the *hybrid degree centrality* of node v, defined as:

$$C_{\text{hybrid}}(v) = (\beta - p)\cdot\alpha\cdot C_{\text{deg}}(v) + p\cdot C_{\text{m-local}}(v), \tag{3}$$

where p is the spreading probability, $C_{\text{m-local}}$ is the modified local centrality, α is a normalizing factor to scale the degree centrality to the magnitude of the modified local centrality, and β is an optimization parameter controlling the ratio between the degree and the modified local centrality. For the datasets considered in [95], the values of $\alpha = 1000$ and β near 0.1 seemed to return favorable results. The modified local centrality is defined as $C_{\text{m-local}}(v) = \sum_{u\in N(v)}\sum_{w\in N(u)} d_2(w) - 2\sum_{u\in N(v)} d(u)$. This modification removes the nearest neighbor influence from the local centrality measure, which is presumably captured somewhat by the degree centrality.

4) VOLUME CENTRALITY

If the spreading process dies out, has a limited reach from its initial source, or has a time out component, then it makes sense that this might be entirely captured by the topology in the local neighborhood of the source node. Kim and Yoneki [96] defined the volume centrality of the node for a given *radius* h as:

$$C_{\text{volume}}(v) = \sum_{u\in N_h(v)} d(u). \tag{4}$$

This is actually a slight modification of the original definition by Wehmuth and Ziviani [97] that uses the set $\tilde{N}_h(v) = N_h(v)\cup\{v\}$. With this latter definition, then when $h = 0$, volume centrality is degree centrality. However, this is already captured when calculating the degrees of nodes in $N_1(v)$. Kim and Yoneki showed that larger h correlates well with closeness centrality (see Eq. (32)). However, as h increases, the complexity of the method will increase. Wehmuth and Ziviani demonstrated that $h = 2$ results in a good trade-off between identifying nodes that diffuse information well and the cost of calculating volume centrality.

5) CLUSTERING COEFFICIENT

One of the characterizations of small-world networks is the increased likelihood of neighbors of a node to be connected. Social networks tend to exhibit this property and an early characterization of this high clustering property is the density

[1]In this instance, we shall abuse our notation by having two representations of degree of the node v, i.e., $C_{\text{deg}}(v) = d(v)$.

of an ego network (i.e., as described in [98], the network of the neighbors of a given node excluding that node). Watts and Strogatz [99] proposed the same metric independently as a way to quantify the clustering of nodes in a given graph and characterize the position of the graph within the spectrum of random to small-world graphs. Their definition has proven incredibly popular. It is expressed by:

$$C_{\text{clustering}}(v) = \frac{1}{d(v) \cdot (d(v)-1)} \sum_{r,s \in N(v)} A_{rs}. \tag{5}$$

Note that each edge will be counted twice in an undirected graph in the summation and the number of such unique edges is normalized by $\binom{d(v)}{2}$, which is the number of possible edges between the neighbors of v. For a directed network, there are twice as many possible directed edges as the undirected case since the adjacency matrix is no longer symmetric, i.e., A_{rs} may not equal A_{sr} and the set $N^{\text{out}}(v)$ of neighbors v links to is used. This measure is often called the *local clustering coefficient* to distinguish it from a global measure of transitivity.

6) REDUNDANCY

Burt [98] introduced the notion of *redundancy* in social networks to describe the concept of neighborhood overlap of a node and its neighbors within the node's ego network. Burt demonstrated redundancy's detriment to *social capital* within socio-economic networks. This is defined as:

$$C_{\text{redundancy}}(v) = \sum_{r \in N(v)} \sum_{s \in N(r) \cap N(v)} p_{vs} m_{rs}, \tag{6}$$

where

$$p_{vs} = \frac{A_{vs} + A_{sv}}{\sum_{r \in N(v)} A_{vr} + A_{rv}},$$

$$m_{rs} = \frac{A_{rs} + A_{sr}}{\max_{t \in N(r) \cap N(v)} A_{rt} + A_{tr}}.$$

Burt uses redundancy to calculate the *effective size (or degree)* of a node's ego (or neighborhood) network taking redundancy into account as $C_{\text{degree}}(v) - C_{\text{redundancy}}(v)$. Borgatti [57] reformulated these expressions to show that for a simple undirected graph, the redundancy is simply $C_{\text{redundancy}}(v) = 2e/C_{\text{degree}}(v)$, where e is the number of links between the neighbors of v, and the effective size of v is $C_{\text{degree}}(v) - 2e/C_{\text{degree}}(v)$. The effective size can be expressed in terms of the degree and clustering coefficient, i.e., $C_{\text{degree}}(v) - (C_{\text{degree}}(v) - 1) \cdot C_{\text{clustering}}(v)$.

7) ENTROPY-BASED MEASURES

In the thermodynamics context, entropy is a measure of the disorder of systems. In the information theory context, entropy measures the amount of information absent in a given process. These concepts of entropy have been used in networks, either in characterizing systems or processes [100]. Nie *et al.* [101] adapted the concept of entropy to centrality. They constructed two variants to measure the entropy, *local entropy* as the contribution of the node to network entropy and *mapping entropy* to incorporate a consideration of the neighbors of the node, defined by:

$$C_{\text{local-entropy}}(v) = -\sum_{u \in N(v)} d(u) \log d(u),$$

$$C_{\text{mapping-entropy}}(v) = -d(v) \sum_{u \in N(v)} \log d(u). \tag{7}$$

8) CLUSTERRANK

As noted with the redundancy measure, high clustering can have an adverse effect on information propagation or spread. With this insight, Chen *et al.* [102] proposed *ClusterRank*, incorporating both the degree as well as the interactions among the neighbors via the clustering coefficient [99]. The ClusterRank of node v is defined as:

$$C_{\text{clusterrank}}(v) = f(C_{\text{clustering}}(v)) \sum_{u \in N^{\text{out}}(v)} (C_{\text{out-deg}}(u) + 1), \tag{8}$$

where Chen *et al.* [102] choose the function $f(x) = 10^{-x}$ and $C_{\text{clustering}}(v)$ refers to the local clustering coefficient defined for directed networks. The summation also adds the out-degree of the node v in the unity term. The coefficient acts as a damping weight where higher clustering is penalized for having fewer unique links to different parts of the network. This damping weight is mitigated if many of the neighbors of v have large numbers of additional neighbors.

9) H-INDEX

Hirsch [92] introduced the *h-index* to measure the impact of the scientific output of a researcher. A researcher has index h if h is the largest integer ℓ such that the researcher has at least ℓ papers each having at least ℓ citations. Korn *et al.* [103] adapted h-index (calling it the *lobby index*) to discover important nodes in networks. A node has index h if the node has at least h neighbors, each having at least degree h, with the rest of the neighbors having at most degree h. Extending this concept, Lü *et al.* [104] defined the $\mathcal{H}$ operator that, for any node v, takes the degrees of the set of its neighbors as an input and returns the maximum number h such that h inputs have value at least h. This can be expressed as:

$$C_{\text{h-index}}(v) = h(v) = \mathcal{H}\left(d(u_1), d(u_2), \ldots, d(u_{d(v)})\right), \tag{9}$$

where $u_1, u_2, \ldots, u_{d(v)} \in N(v)$. If the zero-order h-index of node v is its degree, i.e., $h^{(0)}(v) = d(v)$, then the value in Eq. (9) can be called the first-order h-index. Then the k-order h-index is defined as $h^{(k)}(v) = \mathcal{H}\left(h^{(k-1)}(u_1), h^{(k-1)}(u_2), \ldots, h^{(k-1)}(u_{d(v)})\right)$; this sequence converges to the coreness, discussed in Section III-B1, as the order increases, i.e., $c_i = \lim_{k \to \infty} h^{(k)}(v)$.

10) CURVATURE

The success of hyperbolic models for networks [105], [106] in reproducing observations from real networks has spurred some interest in measuring the intrinsic geometry of complex networks. Curvature in networks is a particularly interesting

aspect to measure since the models typically presume a constant curvature but the reality (and data) is rarely that convenient. There are several competing approaches for curvature. Eckmann and Moses [107] derived a curvature formulation that is equivalent to the local clustering coefficient of Watts and Strogatz [99] and is used to reveal a connection between high curvature and common topics in the World Wide Web. A popular approach is derived from a Gaussian curvature on planar graphs [108], [109], that has been generalized for complex networks [110] as:

$$C_{\text{Gauss-curv}}(v) = \sum_{k\geq 0}(-1)^k \frac{s_v^{k+1}}{k+1}, \tag{10}$$

where s_v^k is the number of k-cliques incident to v. A truncated version of this is used in [111] to compare a network model with data. A third approach of recent interest adapts a notion of Ricci curvature to networks via the transfer of a mass distribution from one vertex to another, and hence can be defined on an edge [112], [113] as $\kappa(u, v) = 1 - W(m_u, m_v)$ and $W(m_u, m_v)$ is the optimal mass transport cost and the mass is typically a unit weight distributed proportionally by an edge weight to the neighbors of the vertices. The curvature at a vertex v is then a weighted sum of the curvature of the incident edges to v, $C_{\text{Ricci-curv}}(v) = \frac{1}{C_{\text{deg}}(v)} \sum_{u\in N(v)} \kappa(u, v)$. Curvature has been shown to have relevance to network fragility [114] and network congestion [115]. An alternate adaptation of Ricci curvature [116], [117] based on the degrees of the vertices has also received some interest due to its simplicity. One very simple version of this Forman-Ricci curvature can be expressed for an edge as $\kappa(u, v) = 2 - C_{\text{deg}}(u) - C_{\text{deg}}(v)$ and for a node as a weighted sum of the curvature of the incident edges.

11) ASYMPTOTIC COMPLEXITY OF LOCAL CENTRALITY METRICS

We summarized the asymptotic complexity using big-O notation of the local centrality metrics surveyed in this section in Table 1. We observe that the centrality of a node is derived from its influence in a given network in terms of influence (e.g., impact or popularity), resourcefulness (e.g., usefulness like social capital), or relationships between neighbors of a node. Local centrality metrics are generally the most easily understood and easiest to compute (with the exception of curvature). The number of neighbors of a node (degree) is inherent to the definition of what a network or graph is. Aggregations of this simple notion, extensions of the neighbors-of-neighbors concept (e.g., semi-local and volume) or aggregations of functions of degree (e.g., entropy and H-index) are also prevalent among local metrics. Another inherent network feature is the network density (fraction of the number of edges to the number of possible edges); these are captured locally by the density of the network of neighbors of a node (e.g., clustering coefficient, ClusterRank, and redundancy). The complexity of many of the metrics (e.g., semi-local, hybrid degree, volume, clustering coefficient, and redundancy) depend on the mean degree in the network. This is because these metrics typically measure the influence of the neighbors of the node to compute the centrality. In the worst case, when the network is dense, the mean degree can be on the order of the network size and the complexity of calculation of many of these metrics is cubic with respect to the network size. In many real networks, i.e., in practice, the degree distribution is such that the vast majority of nodes have small degree and the computation of their local centrality is significantly easier. The 'local' in local centrality implies that the extent of the measure is limited to its neighborhood. While the relevance of a local centrality is derived from this inherent nature, their simplicity can also fail to capture more complex interactions over long time periods (e.g., information spreading and epidemics).

B. MEASURES OF ITERATIVE CENTRALITY

Iterative centrality metrics rely on iterative processes to calculate the centrality value of each node. In some cases, the number of iterations is fixed and determined by a characteristic of the network (e.g., maximum degree), and these metrics still incorporate mostly local information of the network. In most cases, the number of iterations depends on the convergence rate of values at each node. Global information is incorporated into the metric at the node via these iterative processes.

1) k-SHELL INDEX OR CORENESS

The most efficient spreaders have been found to reside in the *core* of the network [118], which can be determined by the process of assigning each node an index (or a positive integer) value derived from the k-shell decomposition. The decomposition and assignment work as follows: Nodes with degree $k = 1$ are successively removed from the network until all remaining nodes have a degree strictly greater than 1. All the removed nodes at this stage are assigned to be part of the k-shell of the network with index $k_S = 1$ or the 1-shell. This is repeated with the increment of k to assign each node to distinct k-shells. The k-shell index of node v is:

$$C_{\text{k-shell}}(v) = \max\{k | v \in H_k \subset G\}, \tag{11}$$

where H_k is the maximal subgraph of G with all nodes having degree at least k in H_k. The coreness and k-shell of networks have been used to characterize network structure, determine network degeneracy, and identify clusters [83].

2) MIXED DEGREE DECOMPOSITION

Zeng and Zhang [119] sought to increase the monotonicity (or the ordering with minimal ties) of the ranking of nodes within each k-shell. They developed a *mixed degree decomposition* that retains elements of the degree mixed with the k-shell index. For node v, this metric is given by:

$$C_{\text{mixed-deg}}(v) = k^{(r)}(v) + \lambda \cdot k^{(e)}(v), \tag{12}$$

where each node starts with mixed degree equal to the residual degree $k^{(r)}(v)$ (i.e., the k-shell index) and the nodes

TABLE 1. Meanings, computations, and complexities of local point centrality metrics.

Centrality name	Meaning	Eq. No.	Complexity	Ref. No.
Degree	Popularity	(1)	$O(n+m)$ or $O(n^2)$	[82], [83]
Semi-local	A node's popularity + the popularity of the node's neighbors	(2)	$O(n\langle k\rangle^2)$	[94]
Hybrid degree	A mixture of degree and a modified semi-local centralities	(3)	$O(n\langle k\rangle^2)$	[95]
Volume	The size of a ball of radius h centered at the node	(4)	$O(n\langle k\rangle^{(h+1)})$	[97], [96]
Clustering coefficient	Probability of node's neighbors being neighbors of each other	(5)	$O(n\langle k\rangle^2)$	[99]
Redundancy	Usefulness (social capital) of a link	(6)	$O(n\langle k\rangle^2)$	[98]
Entropy-based measures	Amount of (missing) information in the node's neighborhood system	(7)	$O(n^2)$	[101]
ClusterRank	Clustering-coefficient weighted semi-local centrality	(8)	$O(nd_{\max}^2+n^2)$	[102]
H-index	Impact (where degree is productivity) of a node's links	(9)	$O(n^2)$	[103]
Curvature	Measure of local geometry near node	(10)	$O(2^n)$	[110]

(Notations: n is the total number of nodes, m is the number of edges, $\langle k\rangle$ is the mean degree of nodes, and $d_{\max}$ is the maximum degree.)

with smallest mixed degrees (M) are removed and assigned to the M-shell. The mixed degrees of the remaining nodes are updated by the current residual degree $k^{(r)}(v)$ and the exhausted degree $k^{(e)}(v)$ (i.e., removed edges from v due to the nodes in the M-shell) and nodes with updated mixed degree not larger than M are also removed and assigned to the M-shell. This is repeated iteratively for the next smallest remaining mixed degree to determine each node's mixed degree. The parameter λ is a value between 0 and 1 that determines the input from the k-shell approach versus the degree approach. When $\lambda = 0$, then mixed degree is simply the k-shell index. On the other hand, when $\lambda = 1$, then mixed degree is simply the degree.

3) NEIGHBORHOOD CORENESS

The core of a network consists of nodes with high k-shell index. Many nodes will have the same high k-shell index. Bae and Kim [120] introduce more diversity by considering the k-shell of neighbors in an approach similar to that of semi-local centrality in Section III-A2. The *neighborhood coreness* and the *extended neighborhood coreness* are defined as:

$$C_{\text{nc}}(v) = \sum_{u\in N(v)} C_{\text{k-shell}}(u) \;\&\; C_{\text{nc+}}(v) = \sum_{u\in N(v)} C_{nc}(u). \quad (13)$$

These metrics introduce more distinguishable monotonicity than using the k-shell.

4) EIGENVECTOR CENTRALITY

This metric is occasionally called Bonacich's degree centrality [27], [55], [87]. Bonacich supported a claim of Cook [35] that centrality is not the same as power and a node with high centrality (e.g., degree) is not necessarily powerful or influential. Accordingly, Bonacich developed an *eigenvector centrality*, which incorporates notions of both centrality and power. The centrality of a node is determined from its direct connections with other nodes and its power is derived from the centralities of these neighbors directly and other nodes in the network indirectly. This latter definition is inspired by the power method (or power iteration), which is an iterative algorithm to attain the principal eigenvalue and eigenvector. The eigenvector centrality of node v is defined as [55]:

$$\begin{aligned} C_{\text{eigenvector}}(v) &= \frac{1}{\lambda}\sum_{u\in N(v)} C_{\text{eigenvector}}(u) \\ &= \frac{1}{\lambda}\sum_{u\in\mathcal{V}} A_{uv}C_{\text{eigenvector}}(u), \end{aligned} \quad (14)$$

where λ is an eigenvalue associated with the principal eigenvector. Note that the iterative approach to attain this centrality requires positive values at initialization to guarantee convergence to the eigenvector corresponding to the maximum eigenvalue, which has non-negative values. Note that the second equality makes clear that the ranking of centralities is determined by the eigenvector of the adjacency matrix.

5) KATZ CENTRALITY

Katz [5] proposed a new status measure by considering the number of direct connections to a node and the statuses of nodes connected to such node. The Katz centrality is defined in [13] as:

$$C_{\text{katz}}(v) = \alpha\sum_{u\in\mathcal{V}} A_{uv}C_{\text{katz}}(u) + \beta, \quad (15)$$

where α is a weight that determines the relative influence of the centrality of the node's neighbors to other nodes in the network by their distances and β is a 'free part' representing a constant extra credit all nodes receive. Katz centrality resolves the problem of zero-valued eigenvector centralities of nodes that do not reside in the strongly connected components of directed graphs [13].

6) AUTHORITY AND HUB CENTRALITIES

Kleinberg [121] introduced an iterative process in the context of hyperlinked web pages to determine which pages are authoritative and which pages are hubs to authoritative pages to assist in web search queries. This approach is often referred to as *Hyperlink-Induced Topic Search (HITS)* in the literature [122]. In this process, each page v is assigned two non-negative weights, one corresponding to its relevance as an authority x_v and another corresponding to its relevance as a hub y_v. Each set of weights are normalized so that the sum

of their squares is unity, i.e., $\sum x_v^2 = 1$ and $\sum y_v^2 = 1$. The update process is given by $x_v \leftarrow \sum_{u:(u,v)\in E} y_u$ and $y_v \leftarrow \sum_{u:(v,u)\in E} x_u$ subject to the normalization invariance. A page's authority depends on the hub weights of the pages linking to it. Similarly, a page's hub weight is determined by the authority weights of the pages it links to. In matrix terms, where $\mathbf{x}$ and $\mathbf{y}$ are vector collections of the authority and hub weights of the nodes, respectively, then the update equations can be expressed as $\boldsymbol{x} \leftarrow \boldsymbol{Ay}/(\boldsymbol{y}^T\boldsymbol{AA}^T\boldsymbol{y})$ and $\boldsymbol{y} \leftarrow \boldsymbol{A}^T\boldsymbol{x}/(\boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{Ax})$. Some simple linear algebra can be used to show that these converge to the principal eigenvectors of the matrices $\boldsymbol{A}^T\boldsymbol{A}$ and $\boldsymbol{AA}^T$, respectively, so long as the initial weights in the process are not orthogonal to the principal eigenvectors. Thus, the authority and hub centrality of the node v is given by:

$$C_{\text{auth}}(v) = [\boldsymbol{e}_1(\boldsymbol{A}^T\boldsymbol{A})]_v, \quad C_{\text{hub}}(v) = [\boldsymbol{e}_1(\boldsymbol{AA}^T)]_v, \quad (16)$$

where $\boldsymbol{e}_1(\cdot)_v$ is the vth element of the principal eigenvector. Kleinberg proposed stopping the process after 10, 000 iterations, as convergence may be slow for large networks.

7) PAGERANK

PageRank is a modern-day variant of Katz centrality that was developed by Brin and Page [123], the founders of Google. PageRank measures the importance of websites by the number of links to the website and is defined in [13] as:

$$C_{\text{pagerank}}(v, \alpha, \beta) = \alpha \sum_{u\in\mathcal{V}, u\neq v} A_{uv} \frac{C_{\text{pagerank}}(u, \alpha, \beta)}{\max(C_{\text{out-deg}}(u), 1)} + \beta, \quad (17)$$

where $C_{\text{out-deg}}(u)$ refers to the out-degree of node u. The interpretations of α and β are similar to the ones described for the Katz centrality in that α is a weight damping the influence of nodes further away from v, while β represents a weight for a free part or credit that each node receives. The key difference is the relative weighting of links to v by the out-degree of the nodes linking to v. In vector form, PageRank can be expressed, with $\beta = 1$, as $\mathbf{C}_{\text{pagerank}}(\alpha, \beta) = (\mathbf{I} - \alpha\mathbf{A}\mathbf{D}^{-1})^{-1}\mathbf{1} = \mathbf{D}(\mathbf{D} - \alpha\mathbf{A})^{-1}\mathbf{1}$, where $\mathbf{D}$ is a diagonal matrix with entries $D_{uu} = \max(C_{\text{out-deg}}(u), 1)$.

8) CONTRIBUTION CENTRALITY

Alvarez-Socorro *et al.* [124] refined the eigenvector centrality to account the similarity of the neighbors that link to a node. The concept presumes that nodes with greater dissimilarity, in the sense of Jaccard [125], should have a greater contribution weight than similar nodes. Nodes that are dissimilar may provide different information than similar nodes. This *contribution centrality* is given by:

$$C_{\text{contribution}}(v) = \frac{1}{\lambda} \sum_{u\in N(v)} W_{u,v} C_{\text{contribution}}(u), \quad (18)$$

where $W_{u,v} = A_{u,v}Dis_{u,v}$ is the contribution of node u to node v and $Dis_{u,v} = 1 - \frac{|N(v)\cap N(u)|}{|N(v)\cup N(u)|}$ is a dissimilarity coefficient. This measure can also be considered as the eigenvector centrality of a weighted network, where the weights are informed by the structural dissimilarity coefficient. The weighted adjacency matrix can be expressed as $\mathbf{W} = \mathbf{A} \odot \mathbf{Dis}$, where $\odot$ is the Hadamard or element-wise product. The λ in Eq. (18) is the maximum eigenvalue of $\mathbf{W}$.

9) DIFFUSION CENTRALITY

Banerjee *et al.* [126] approximate communication centrality (i.e., the fraction of the number of nodes that choose to participate in the purchase of a product after being informed versus the total number of nodes that were informed of the product). Their *diffusion centrality* can be expressed in vector form as:

$$\mathbf{C}_{\text{diffusion}}(q, T) := \sum_{t=1}^{T} (q\mathbf{A})^t \mathbf{1}, \quad (19)$$

where q is the passing probability and T is the number of iterations. This centrality captures a number of different measures depending on the value of T or the number of iterations of information sharing or passing. When $T = 1$, then diffusion centrality will be proportional to degree centrality. When $T \to \infty$, $\mathbf{A}$ is diagonalizable (this is always true for real symmetric matrices, thus true for undirected network adjacency matrices), and $q \geq \frac{1}{\lambda}$ (where λ is the maximum eigenvalue of $\mathbf{A}$), then diffusion centrality is proportional to eigenvector centrality. But when $q < \frac{1}{\lambda}$, this is a type of Katz-Bonacich centrality.

10) SUBGRAPH CENTRALITY

Subgraph centrality measures the weighted sum of the closed paths incident to v in the network, including both cyclic and acyclic paths (i.e., a path that backtracks on itself), where the contribution or weight of each path in the sum decreases as the path length increases [127]. Thus, this metric measures the inclusion of the node in all connected subgraphs of the network and is characterized significantly by the inclusion of the node in motifs or small subgraph patterns. Subgraph centrality is given by:

$$C_{\text{subgraph}}(v) = \sum_{k=0}^{\infty} \frac{\mu_k(v)}{k!} = \sum_{j=1}^{N} (u_j^v)^2 e^{\lambda_j}, \quad (20)$$

where $\mu_k(v) = (\mathbf{A}^k)_{vv}$, λ_j is the jth eigenvalue of $\mathbf{A}$ and u_j is its corresponding eigenvector (u_j^v is the vth element of this vector). Inclusion in smaller subgraphs (closed walks) is given more significance due to the scaling, which is also necessary for convergence of the sum. The measure is useful to distinguish between nodes with equivalent values of degree centrality, betweenness, closeness, or eigenvector centrality. The authors conjecture that if the subgraph centrality is identical for all nodes, then these other measures will also be identical. Note that the average centrality of all the nodes is trivial to determine to be $\langle C_{\text{subgraph}} \rangle = \frac{1}{N}\sum_{i=1}^{N} e^{\lambda_j}$.

11) LEADERRANK

Lü *et al.* [128] proposed *LeaderRank* to find prominent members, or *leaders*, and thereby rank them in terms of their influence, particularly in a social network context. Given a *leadership network* or a directed graph with *leaders* and *fans*, where a directed edge existing signifies the subscription from a fan to a leader, LeaderRank generates a supplemental network, created via the addition of *ground node g* with bidirectional edges between all the nodes in the leadership network. This ensures a strongly connected graph with $n+1$ nodes and $m+2n$ directed edges containing the subgraph of the original leadership network of n nodes and m directed edges. Each node, except the ground node, is assigned an initial unit score. In each unit of time or iteration, the current score of each node is equi-distributed to the neighbors the node is linked to, until equilibrium. The proportion of score allocated from node u to node v in one unit of time is $A_{uv}/C_{\text{out-deg}}(u)$. At time t, the amount of score allocated at node v is $s_v(t+1) = \sum_{u\in N^{\text{in}}(v)}^{n+1} \frac{1}{C_{\text{out-deg}}(u)} s_u(t)$, where $s_v(0) = 1$ for all non-ground nodes and $s_g(0) = 0$. At the equilibrium time t_e, the score of the ground node is equi-distributed to the other nodes, which ensures no loss of value in the distribution scheme for the leadership network. Hence, the final LeaderRank score of node v is:

$$C_{\text{LeaderRank}}(v) = s_v(t_e) + \frac{s_g(t_e)}{n}. \tag{21}$$

12) DYNAMICAL INFLUENCE

Klemm *et al.* [129] proposed the concept of dynamic influence as a centrality measure that can quantify the influence of a node's dynamic state on the collective system behavior based on the interplay between dynamics and structure in complex networks. Given systems with n time-dependent real variables, $\mathbf{x} = [x_1, \cdots, x_N]$ associated with linear dynamics denoted by an $n \times n$ real matrix, $\mathbf{M}$, we have the update function $\dot{\mathbf{x}} = \mathbf{M}\mathbf{x}$. The largest eigenvalue $\mu_{\max}$ for $\mathbf{M}$ is considered to obtain a first classification of dynamics. When $\mu_{\max}$ is negative, $\mathbf{x}(t)$ converges to a null vector as a stable, fixed solution. When $\mu_{\max}$ is positive, $\mathbf{x}(t)$ will grow indefinitely from the initial state $\mathbf{x}(0)$. Assuming that there exists a non-degenerate $\mu_{\max}$ for $\mathbf{M}$, we define a scalar product $\phi_c = \mathbf{c}\cdot\mathbf{x}$ as a conserved quality where $\mathbf{c}$ is a left eigenvector of $\mathbf{M}$ for $\mu_{\max}$ governed by $\frac{d\phi_c}{dt} = \mathbf{c}\cdot\dot{\mathbf{x}}(t) = [\mathbf{cM}]\cdot\mathbf{x}(t) = 0$. When the conserved quality exists, the final state can be calculated from the initial state $\mathbf{x}(0)$ by:

$$\mathbf{C}_{\text{dynamic-influence}} := \mathbf{x}(\infty) = \lim_{t\to\infty} \mathbf{x}(t) = \frac{\mathbf{c}\cdot\mathbf{x}(0)}{\mathbf{c}\cdot\mathbf{e}}\mathbf{e}, \tag{22}$$

where $\mathbf{e}$ refers to a right eigenvector of $\mathbf{M}$ for $\mu_{\max}$. The above equation means that $\mathbf{C}_{\text{dynamic-influence}}$ is projected based on $\mathbf{x}(0)$ where c_i represents the effect of $\mathbf{x}(0)$ on the final state $\mathbf{x}(\infty)$.

13) CUMULATIVE NOMINATION

Poulin *et al.* [130] introduced *cumulative nomination* whereby the reputation of a node is derived from the nominations of its neighbors and, hence, a node located more centrally in the network is nominated more frequently than a node located on the periphery. Initially, a unit of nomination is provided to each node in the network. Then for each nomination round or iteration, the nomination value of each node is updated as the sum of the nominations from its neighbors, i.e., for node v, the nomination update is determined from $p'^n_v = p'^{n-1}_v + \sum_{u\in N(v)} p'^{n-1}_u$, where $p'^0_v = 1$. It is convenient to normalize this process at each step: $p^n_v = \frac{p^{n-1}_v + \sum_{u\in N(v)} p^{n-1}_u}{\sum_{w\in\mathcal{V}}\left[p^{n-1}_w + \sum_{u\in N(w)} p^{n-1}_u\right]}$. At equilibrium, the cumulative nomination of node v is given by:

$$C_{\text{cumulative-nomination}}(v) = \lim_{n\to\infty} p^n_v. \tag{23}$$

This metric is analogous to the one proposed by Bonacich [55], but it is empirically proven to be faster in convergence to the steady state [130].

14) SALSA

Lempel and Moran [131] developed a *Stochastic Approach for Link Structure Analysis* (or *SALSA*) as an alternative to the hubs and authorities approach of [121] for web links. The given directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is converted into an undirected bipartite graph $\tilde{G} = (\tilde{V}, \tilde{E})$ between a hub side $V_h \subset \tilde{V}$ and an authority side $V_a \subset \tilde{V}$. Each node v in $\mathcal{V}$ is represented by two nodes in $\tilde{V}$, one on the hub side v_h and one on the authority side v_a. Each directed edge from v to u in $\mathcal{E}$ is represented by an undirected edge between v_h and u_a in $\tilde{E}$. Two random walks, starting from either side of $\tilde{V}$, of path length two, construct Markov chains that reveal a ranking of nodes as hubs and authorities in the network. The transition matrices of these Markov chains can be defined by a hub matrix $\tilde{\mathbf{H}}$, with element entries $\tilde{H}_{u,v} = \sum_{x\in\mathcal{V}|(u_h,x_a),(v_h,x_a)\in\tilde{E}} \frac{1}{d(u_h)}\cdot\frac{1}{d(x_a)}$, and an authority matrix $\tilde{\mathbf{A}}$, with entries $\tilde{A}_{u,v} = \sum_{x\in\mathcal{V}|(x_h,u_a),(x_h,v_a)\in\tilde{E}} \frac{1}{d(u_a)}\cdot\frac{1}{d(x_h)}$, where the degree is in $\tilde{G}$. The updates for these transition matrices are $\mathbf{h}^n = \tilde{\mathbf{H}}\mathbf{h}^{n-1}$ and $\mathbf{a}^n = \tilde{\mathbf{A}}\mathbf{a}^{n-1}$, where the initial value assigned for each node is 1. As with the mutual reinforcement approach of Kleinberg's hubs and authorities, the principal eigenvectors of the transition matrices are the convergent points of the iterations, i.e.,

$$\begin{aligned} C_{\text{SALSA-hub}}(v) &= [e_1(\tilde{\mathbf{H}})]_v, \text{ and} \\ C_{\text{SALSA-auth}}(v) &= [e_1(\tilde{\mathbf{A}})]_v. \end{aligned} \tag{24}$$

15) ASYMPTOTIC COMPLEXITY OF ITERATIVE CENTRALITY METRICS

We summarized the iterative point centrality metrics, their meanings, and Eqs. numbers, asymptotic complexity using big-O notation, and the corresponding sources in Table 2. The key advantage of iterative centrality metrics lies in their ability to capture influences of nodes beyond their local neighborhoods yet still based on local interactions (calculations). This enables these metrics to capture more complex dynamical relationships since a node's influence is determined by the

TABLE 2. Meanings, computations, and complexities of iterative point centrality metrics.

Centrality name	Meaning	Eq. No.	Complexity	Ref. No.
k-shell index or coreness	Hierarchical structure membership of a node in the network	(11)	$O(n+m)$	[118]
Mixed degree decomposition	Mixture of k-shell and degree	(12)	$O(n+m)$	[119]
Neighborhood coreness	Aggregating k-shell indices of neighboring nodes	(13)	$O(n^2+m)$	[120]
Eigenvector	Importance of neighboring nodes to determine node's importance	(14)	$O(n^3)$	[55]
Katz	Adding damping effect on distant nodes to eigenvector centrality	(15)	$O(n^3)$	[5]
Authorities & Hubs	Eigenvector centrality for directed networks	(16)	$O(n^3)$	[121]
PageRank	Google's algorithm that adapts Katz centrality, weighting influence by out-degree	(17)	$O(n^3)$	[123]
Contribution	Weighted eigenvector centrality using structural dissimilarity	(18)	$O(n^3)$	[124]
Diffusion	Influence of the spread of information over finite time	(19)	$O(n^3)$	[126]
Subgraph	Incidence of nodes to closed walks weighted by length (motifs)	(20)	$O(n^3)$	[127]
LeaderRank	Parameterless PageRank	(21)	$O(n^3)$	[128]
Dynamical influence	Initial dynamic state incorporated into the eigenvector centrality	(22)	$O(n^3)$	[129]
Cumulative nomination	Reputation through the nomination process approaching Bonacich centrality	(23)	$O(n^3)$	[130]
SALSA	Random walk alternative to hubs & authorities	(24)	$O(n^3)$	[131]

(Notations: n is the total number of nodes and m is the number of edges.)

influence of its neighbors. This, of course, requires a number of iterations from a suitable initialization so that the update defining this dependence on the neighbors converges. The most prevalent of these are the eigenvector-type centralities (e.g., eigenvector, Katz, authorities and hubs, pagerank, salsa, subgraph, etc.). These centralities are essentially solutions to the primary eigenvector for some weighted variant of the adjacency matrix. Other iterative approaches capture more structural features (e.g., k-shell, neighborhood coreness, etc.) with a finite number of iterations determined by the maximum degree. This set of centralities are relevant for notions of influence since they capture nodes that have many neighbors that themselves have many neighbors. Regarding the asymptotic complexity of these iterative point centrality metrics, there are several trends distinct from the local point centrality metrics. Generally, but certainly for sparse networks, the eigenvector-based iterative centralities are more computationally expensive to compute. This is because the mean degree that is prominent in many local centrality complexities typically satisfies $\langle k\rangle \ll n$. However, the k-shell-based iterative centralities can be computationally cheap to compute. This is because the number of iterations is fixed based on the maximum degree (and for some metrics the size of their k-shell). One approach that helps ease the computational burden of the eigenvector-based iterative centralities is suggested in the solution used for the authorities and hubs centrality (or HITS algorithm). This is to stop the update before convergence after some large, but fixed number of iterations. This still ensures that the influence of nodes beyond the local neighborhood and can be treated as an approximation of the desired centrality determined by the given update function.

C. MEASURES OF GLOBAL CENTRALITY

Global centrality metrics require a measurement using possibly the entire network topology. These approaches involve the measurement of path lengths between nodes that are separated (non-adjacent) in the network. The calculations of shortest paths often do not scale well with network size; hence, these metrics are generally more computationally expensive. We describe 17 graph centrality metrics in the following sections.

1) IMPROVED METHOD OR k-SHELL DISTANCE

Liu *et al.* [132] introduced an *improved method* or *k-shell distance* as an alternative approach to distinguish these intra-k-shell nodes, whereby each node in the k_S core is further ranked by $\theta(v|k_S) = (k_S^{\max} - k_S + 1)\sum_{u\in J} d(v,u)$, where $k_S^{\max}$ is the largest k-shell index in the network and J is the network core (nodes in the subset with the largest k-shell index). This centrality can be considered as a two element vector:

$$C_{\text{improved-method}}(v) = (k_s, \theta(v|k_s)), \tag{25}$$

where nodes are sorted first by k_s from largest to smallest and then, for the same k_s, by $\theta(v|k_s)$ from smallest to largest. Essentially, nodes within the same k-shell are distinguished by how close the nodes are to all other nodes in the network core.

2) BETWEENNESS CENTRALITY

One early concept of centrality capturing the notion of betweenness, learned from studies on human interactions in a laboratory setting [2], [4], was derived from the observation of certain nodes having control on the communication between a pair of other nodes based on their position in the network. The ability of a node to control this communication grants it a position of influence as a broker or enabler. Locally, a node with high degree has potential for fulfilling such a role, depending on the level of clustering (links) between the neighbors of the node, but this would be true only for its immediate neighbors. It does not capture the control the

node has on the communication between a pair of nodes that are distant from each other. This concept was finally encapsulated as *betweenness centrality* in [39]. For a node v, its betweenness is given by:

$$C_{\text{bet}}(v) = \sum_{s,t|s\neq v\neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}, \tag{26}$$

where σ_{st} is the number of shortest paths between s and t and $\sigma_{st}(v)$ is the number of shortest paths between s and t that include v. For comparing the relative betweenness between nodes in different networks, the centrality can be scaled or normalized [82] by $\binom{n-1}{2}$, which is the number of possible pairs of shortest paths node v can be between. This extreme example only occurs for the center node in a star network. Betweenness centrality has received significant interest in applications in information flow [133], network resilience [15], or network classification [134]. A variant of this centrality adapted for edges is popularly used to detect community structure [135]. This interest has led to efforts to improve the computation of betweenness [59], although for large and dense networks, the measure can still become computationally prohibitive.

3) L-BETWEENNESS CENTRALITY

Ercsey-Ravasz and Toroczkai [136] formalized a notion of betweenness, called *L-betweenness*, to reduce the computational costs of its calculation. This centrality originally described by [56], considering shortest paths of length at most L, i.e.,

$$C_{\text{L-bet}}(v) = \sum_{s,t|s\neq v\neq t, d(s,t)\leq L} \frac{\sigma_{st}(v)}{\sigma_{st}}. \tag{27}$$

If L is at least the diameter of the network, then L-betweenness is equivalent to betweenness centrality. Ercsey-Ravasz and Toroczkai [136] explicitly express this quantity in terms of the summation of betweenness centralities at each vertex for shortest paths of fixed length ℓ over the range $\ell = 1, \ldots, L$. That construction is particularly useful for their analysis demonstrating a scaling factor with respect to L and that for relatively small values of L, the L-betweenness centrality is a good indicator of the true betweenness centrality in terms of ranking the nodes. For small L, this metric straddles the boundary between the classes of global and local centrality metrics.

4) FLOW BETWEENNESS CENTRALITY

Freeman *et al.* [60] proposed a variant of betweenness to capture the capacity of information that can flow in a valued or weighted graph. The concept borrows from maximum flow-minimum cut theory [137]. Given the maximum flow m_{rs} between vertices r and s, denote by $m_{rs}(v)$ the portion of this flow that passes through node v. Then the *flow betweenness* for node v is given by:

$$C_{\text{flow-bet}}(v) = \sum_{s,t|s\neq v\neq t} m_{st}(v). \tag{28}$$

This expression can be normalized by replacing each summand $m_{st}(v)$ with $\frac{m_{st}(v)}{m_{st}}$. This metric can be used to estimate the mean difference between the highest centrality and the centralities of the other nodes as a graph centrality metric, as discussed in Section IV-A4.

5) RANDOM-WALK BETWEENNESS CENTRALITY

Like flow betweenness, this also captures a notion of betweenness beyond shortest paths. Newman [138] introduced *random-walk betweenness* to incorporate the contribution from all paths (short and long) with more weights given to shorter paths. Actually, Newman first defined the measure via a current flow analogy and showed it is equivalent to random walks. Formally, this measure is defined by:

$$C_{\text{random-walk-bet}}(v) = \frac{\sum_{s,t|s<t} I_v^{(st)}}{\frac{1}{2}n(n-1)}, \tag{29}$$

where $I_v^{(st)} = \frac{1}{2}\sum_u A_{vu}|T_{vs} - T_{vt} - T_{us} + T_{ut}|$ and T is the matrix $(D_w - A_w)^{-1}$ where $D_w - A_w$ is the Laplacian with the w-th row and column removed (e.g., the last column and row). Note $I_s^{(st)} = I_t^{(st)} = 1$.

6) LOAD CENTRALITY

In the context of the transportation of data over a network, high centrality nodes encounter a heavy load in terms of the data packets that may be transmitted over shortest paths. Goh *et al.* [139] defined the load centrality of node v as the total quantity of data packets traversing over node v after every node in the network sends a single packet to every other node along a shortest path. For the scenario where more than one shortest path exists between two nodes, the quantity is divided at each branching point evenly. Explicitly,

$$C_{\text{load}}(v) = \sum_{s,t|s\neq v\neq t} \theta_{st}(v), \tag{30}$$

where $\theta_{st}(v)$ is the amount of the unit quantity that passed through node v from node s to node t such that the quantity is split uniformly at each branch encountered in a shortest paths from s to t. There has been some confusion that this load centrality is equivalent to the betweenness centrality (even in the original paper [139]). However, the quantity in betweenness is split evenly along each shortest path and not at the branching points. For this reason, it is often the case that even in simple graphs the load due to a pair of vertices is not symmetric at every vertex, i.e., $\theta_{st}(v) \neq \theta_{ts}(v)$. A simple algorithm for the calculation of load is provided in [140].

7) ROUTING BETWEENNESS CENTRALITY

Dolev *et al.* [141] defined a variant of betweenness based on the particular routing strategy that considers the effect of the potential traffic load on the network. This *routing betweenness centrality* measures the expected number of packets passing through a given vertex. For the vertex v, the routing

betweenness is calculated by:

$$C_{\text{routing-bet}}(v) = \sum_{s,t \in \mathcal{V}} \sigma_{st}(v) \cdot T(s,t), \tag{31}$$

where $\sigma_{st}(v)$ is the probability that a packet will go through v when it is sent from s to t, and $T(s,t)$ is the total number of paths from s to t. This probability is dependent on the particular routing protocol.

8) CLOSENESS CENTRALITY

Bavelas [3] was interested in distinguishing between different positions in small group networks. His approach was *closeness centrality*, defined as the reciprocal of *farness*, or the inverse proportion of the average distance to all other nodes in the network. Formally, this can be expressed as:

$$C_{\text{closeness}}(v) = \frac{1}{\sum_{u \in \mathcal{V}} d(v,u)}. \tag{32}$$

Often, this quantity is normalized for comparisons across networks by multiplying by $n-1$ (or n for large networks). Another approach to compare the relative position of nodes with the same farness in different structure groups is given by [3], $C_{\text{bavelas}}(v) = \frac{\sum_{s,t \in \mathcal{V}} d(s,t)}{\sum_{u \in \mathcal{V}} d(v,u)}$, which is equivalent to $C_{\text{closeness}}(v) / \sum_{u \in \mathcal{V}} C_{\text{closeness}}(u)$.

9) INFORMATION CENTRALITY

Stephenson and Zelen [26] developed a centrality measure that uses all paths between pairs of nodes to incorporate the notion of the potential transmission of information. This *information centrality* borrows from the statistical estimation perspective that there is noise from a signal transmission captured by the variance of the signal passing through a path so that the information decreases as the distance between nodes grows. Treating this variance as unity for each link, the information for node v is then defined as the harmonic mean of the information between v and every other node, that is,

$$C_{\text{information}}(v) = \frac{n}{\sum_{u \in \mathcal{V}} \frac{1}{I_{uv}}}, \tag{33}$$

where I_{uv} is the information along all paths from u to v, weighted by the length of each path. This quantity is ultimately given by $I_{uv} = 1/(C_{uu} + C_{vv} - 2\, C_{uv})$, where $\mathbf{C} = \mathbf{D} - \mathbf{A} + \mathbf{1}\mathbf{1}^T$, $\mathbf{D}$ is a diagonal matrix of node degrees and $\mathbf{1}$ is a vector of ones. Hence, the information centrality can be rewritten as $C_{\text{information}}(v) = (C_{vv} + \frac{\text{tr}(\mathbf{C})}{n} - \frac{2}{n^2})^{-1}$.

10) CURRENT-FLOW BETWEENNESS AND CLOSENESS

An alternative notion of flow, similar to the max-flow-min-cut approach for flow betweenness, is to model information spread over a network as an electric current [142]. *Current-flow betweenness* is defined as:

$$C_{\text{current-bet}}(v) = \frac{1}{(n-1)(n-2)} \sum_{s,t \in \mathcal{V}} \tau_{st}(v), \tag{34}$$

where $\tau_{st}(v)$ is the electrical current that passes through node v given a supply entering the source node s and exiting the terminus node t. More formally, $\tau_{st}(v) = \frac{1}{2}\left(-|b(v)| + \sum_{e:v \in e} |x(\overrightarrow{e})|\right)$. Here, $\mathbf{b}$ is a vector defining the current supply, where current enters and exits at nodes s and t, respectively, i.e., $b(s) = 1$, $b(t) = -1$, and b is zero elsewhere. The current over each edge in the network is given by the vector $\mathbf{x}$ such that it satisfies Kirchhoff's Current and Potential Laws. This is equivalent to *random-walk betweenness* [138]. This approach with current can be extended to other path-based centralities. For example, *current-flow closeness* is defined as:

$$C_{\text{current-closeness}}(v) = \frac{n-1}{\sum_{w \neq v} p_{vw}(w) - p_{vw}(w)}, \tag{35}$$

where $p(\overrightarrow{e}) = x(\overrightarrow{e})/c(e)$ by Ohm's Law, and where the conductance $c(e)$ is the inverse of the resistance $r(e)$ or length of an edge. This variant of closeness has been shown to be equivalent to *information centrality* [26].

11) RESIDUAL CLOSENESS

Dangalchev [143] developed *residual closeness* to determine the vulnerability of the graph (of becoming disconnected with the removal of a few nodes or edges) using a variation of closeness. This is defined by:

$$C_{\text{residual-closeness}}(v) = \sum_{u \neq v} \left(\frac{1}{2}\right)^{d(v,u)}. \tag{36}$$

Rather than taking the reciprocal of the sum of distances, residual closeness uses a weighting scheme. A generalization of this idea already exists in the literature [144], although it was not explicitly expressed as a centrality metric until later [145]. Jackson [145] calls this metric *decay centrality*, expressed as $C_{\text{decay}}(v) = \sum_{u \neq v} \delta^{d(v,u)}$, where $0 < \delta < 1$ is the decay rate. Recently, Tsakas [146] has shown that the maximum decay centrality often coincides with the maximum degree centrality when $\delta > \frac{1}{2}$ and with the maximum closeness centrality when $\delta < \frac{1}{2}$, at least on Erdös-Rényi graphs.

12) STRAIGHTNESS CENTRALITY

In spatial networks, the distance between neighbors is not uniform (or unweighted). Crucitti *et al.* [147] applied and developed generalizations of some common metrics that account for the network's embedding in space. Closeness and betweenness centralities are identical to their weighted distance versions [83], i.e., the distance between two nodes is the true distance (or weight) from one node to the other. The new metric developed by [147] is *straightness centrality*, which is given for node v by:

$$C_{\text{straightness}}(v) = \frac{1}{n-1} \sum_{u \in \mathcal{V}, u \neq v} \frac{d_{\text{Euclidean}}(u,v)}{d(u,v)}, \tag{37}$$

where $d_{\text{Euclidean}}(u,v)$ is the Euclidean distance in the real or embedded space. Straightness centrality measures the efficiency of the route between two nodes using node v.

13) AHP-BASED CENTRALITY

Bian *et al.* [148] developed the *Analytic Hierarchy Process (AHP)* as a decision making process adapted to identify influential nodes. The steps to compute it are as follows:

1) Calculate centrality values (e.g., degree, betweenness, closeness) for each node and combine in an $n \times 3$ matrix.
2) Calculate weights. Bian *et al.* [148] appended another vector to the above matrix derived from results of a Susceptible-Infected (SI) process run on the nodes [149]. This produces the matrix $\mathbf{D} = [\mathbf{C_D}, \mathbf{C_B}, \mathbf{C_C}, \mathbf{F(t)}]$, where $\mathbf{D}$ is an $n \times 4$ matrix, $\mathbf{C_D}$, $\mathbf{C_B}$, and $\mathbf{C_C}$ are vectors for degree, betweenness, and closeness centralities, respectively, and $\mathbf{F(t)}$ is a vector of results of the SI process. The matrix is normalized and weights are determined by matching the attributes to the SI column, i.e., $r_{ij} = \frac{D_{ij}}{\sum_{i=1}^{n} D_{ij}}$, for $i = 1, \ldots, n; j = 1, \ldots, 4$, $v_{ij} = \frac{1}{|r_{ij} - r_{i4}|}$ for $i = 1, \ldots, n; j = 1, 2, 3$, $e_j = \sum_{i=1}^{n} v_{ij}$, and finally $w_j = \frac{e_j}{\sum_{j=1}^{3} e_j}$, for $j = 1, 2, 3$. $\mathbf{w}$ is a 3×1 vector that represent the weight for three metrics.
3) Calculate the matrix of option scores using the AHP, i.e., $B_{ik}^{(j)} = \frac{D_{ij}}{D_{kj}}$ for $i = 1, \ldots, n; k = 1, \ldots, n; j = 1, 2, 3$, where $\mathbf{B}^{(j)}$ is an $n \times n$ matrix. Then the option scores are $\mathbf{s_j} = \lambda \times \mathbf{B}^{(j)}$, for $j = 1, 2, 3$, where λ is the largest eigenvalue of matrix $\mathbf{B}^{(j)}$.
4) The nodes are then ranks using the so called Analytical Hierarchy Process (APT) by:

$$\mathbf{C}_{\text{AHP}} = \mathbf{s} \times \mathbf{w}^T, \tag{38}$$

where $\mathbf{s}$ is $n \times 3$ matrix with columns $\mathbf{s}_j$ for $j = 1, 2, 3$ and $\mathbf{w}^T$ is a transpose vector of $\mathbf{w}$, which is a vector of weights w_j for $j = 1, 2, 3$, respectively.

The presumption is that the SI scores in the above process are based on short time horizons, whereas the results of the AHP may have value for longer time horizons. Thus, AHP combines three classic centrality metrics and weights them via a short-run epidemic compartmental model process.

14) GENERALIZED DEGREE AND SHORTEST PATHS (GDSP)

For weighted networks, extensions to the usual centrality measures already exist for degree [150], closeness [151], and betweenness [59]. In incorporating weights, the measures ignore the number of ties or intermediaries. Opsahl *et al.* [152] sought to remedy this with the creation of generalized measures that also encompass both the traditional measures and the weighted versions:

$$\begin{aligned} C^{w}_{\text{gen-deg}}(v, \alpha) &= C_{\text{deg}}(v)^{(1-\alpha)} \cdot C^{w}_{\text{deg}}(v)^{\alpha}, \\ C^{w}_{\text{gen-closeness}}(v, \alpha) &= \left[\sum_{u} d^{w}(v, u, \alpha) \right]^{-1}, \\ C^{w}_{\text{gen-bet}}(v, \alpha) &= \sum_{s,t} \frac{\sigma^{w}_{st}(v, \alpha)}{\sigma^{w}_{st}}, \end{aligned} \tag{39}$$

where the shortest path weighted distances given by $d^{w}(u, v) = \min\left(\frac{1}{w_{ui_1}} + \cdots + \frac{1}{w_{i_k v}}\right)$ are replaced with $d^{w}(u, v, \alpha) = \min\left(\frac{1}{(w_{ui_1})^{\alpha}} + \cdots + \frac{1}{(w_{i_k v})^{\alpha}}\right)$ for the shortest path $u, i_1, \ldots, i_k, v$. For each generalization, α is a tuning parameter depending on the research or application setting. When $\alpha = 0$, the measures are the usual (unweighted) centrality measures; when $\alpha = 1$, the measures are the common weighted measures. When $\alpha \in (0, 1)$, having many weak ties correlates with higher generalized centrality; and when $\alpha > 1$, having fewer weak ties correlates with higher generalized centrality.

15) WEIGHT NEIGHBORHOOD CENTRALITY

Wang *et al.* [153] included a notion of the diffusion importance of links based on the power-law property found in the distribution of observed centrality values (e.g., degree, betweenness) in real networks. Their *weight neighborhood centrality* is defined as:

$$C_{\text{weight-nbhd}}(v, \phi) = \phi_v + \sum_{u \in N(v)} \frac{w_{uv}}{\langle w \rangle} \cdot \phi_u, \tag{40}$$

where the weights are given by $w_{uv} = (C_{\text{deg}}(u) \cdot C_{\text{deg}}(v))^{\alpha}$ and ϕ is the benchmark centrality (e.g., degree, betweenness, k-shell), α is a tunable parameter between 0 and 1, and $\langle w \rangle$ is average weight for edges.

16) PERCOLATION CENTRALITY

Piraveenan *et al.* [154] developed *percolation centrality* to capture the dynamic changes of a network topology based on the percolation process. Typically, the percolation state of a node v at time t might be denoted by $x_v(t)$ and has a discrete value corresponding to current state, either percolated (on) or not percolated (off), where a 0 value indicates v is not percolated (e.g., infected) at time t and a value of 1 indicates it is percolated. When $0 < x_v(t) < 1$, then v might be said to be is in the process (or probability) of being percolated. Hence, a higher value of $x_v(t)$ implies that v is closer to (has a greater chance of) being percolated. Piraveenan *et al.* [154] defined this percolation centrality as the proportion of percolated paths passing through a node, which for node v is measured by:

$$\begin{aligned} &C_{\text{percolation}}(v, t) \\ &\quad = \frac{1}{n-2} \sum_{r \neq v \neq s} \frac{\sigma_{rs}(v)}{\sigma_{rs}} \frac{x_r(t)}{[\sum_{u \in \mathcal{V}} x_u(t)] - x_v(t)}, \end{aligned} \tag{41}$$

where σ_{rs} is the total number of shortest paths between r and s and $\sigma_{rs}(v)$ is the total number of shortest paths between r and s passing through v. When only a single source node is (partially) percolated, then the average of the percolation centrality for every node over all possible sources (excluding itself) is proportional to betweenness centrality (see Eq. (26)) as $x_r(t)/([\sum_{u \in \mathcal{V}} x_u(v)] - x_v(t)) = 1$ when only when x_r is the source, thereby contributing a $1/(n-1)$ factor. If all nodes are (partially) percolated at the same level, all shortest paths are percolated paths, leading to the state that percolation centrality is proportional to betweenness centrality.

17) ECCENTRICITY

Based on the idea that the centrality of a node depends on the distance, i.e., the shortest path, between other nodes in networks, Per and Frank [155] introduced the concept of *eccentricity*, which is the maximum distance between a node and any other node in the network. Lower eccentricity indicates higher centrality. Eccentricity centrality can be mathematically expressed as:

$$C_{\text{eccentricity}}(v) = \frac{1}{\max\{d(v,u)|u \in V\}}. \tag{42}$$

18) ASYMPTOTIC COMPLEXITY OF GLOBAL CENTRALITY METRICS

In Table 3, we summarized the meanings, equations, asymptotic complexity in big-O notation of the global point centrality metrics surveyed in this section. Global centrality metrics have the benefit of capturing the global connectivity of the entire network. These centralities are typically derived by notions of distance between all pairs of vertices in the network (e.g., betweenness, load, closeness, etc.). As such, they provide more information about clustering between parts of a network. Global centralities help identify the key nodes that interact between different parts of the network revealing very different features than notions of degree (e.g., community structure, brokerage, and control). Unfortunately, these metrics are also significantly more burdensome to calculate due to the nontrivial task of determining the distance or shortest path for each node pair. Every distance or path must be found to determine its contribution to the value for every node. This dependence on calculating paths is reflected in Table 3, where we can observe that the complexity for many of the centralities is the function of the number of edges m. This also means that the denser the network becomes the more costly the computation for these metrics. This is, in part, because there are more potential paths to consider in finding the shortest paths. Comparing to the iterative point centrality metrics in Table 2, we observe that the global centrality metrics are more costly, since the number of edges is at least the order of n for the network to be connected. The ideal approach to limit this complexity cost is restricting the length of the paths considered, as done in the L-betweenness metric. The tradeoff is having to learn what the appropriate maximum length L should be considered to achieve a good approximation of the desired centrality while limiting the cost significantly enough to make the exercise worthwhile.

D. DISCUSSIONS: POINT CENTRALITY METRICS

In Section III, we extensively surveyed three types of point centrality metrics: *local*, *iterative*, and *global*. We discussed the meaning of each metric and described how to compute it. In addition, we analyzed their asymptotic complexity in big-O in order to discuss how efficient each metric is and to have better insights on how to utilize them under different environmental conditions in terms of network and node density, network dynamics, the presence of attacks, or resource constraints. Since the key aspect of point centrality metrics is to measure a node's centrality in terms of its power, influence, or importance, depending on the size of a given network, computational overhead to estimate the node's centrality is heavily affected. Particularly under resource-constrained environments or when a given network's topology is only partially observable, we can think of how to determine a node's ego network which can be considered to determine its centrality, rather than assuming that all nodes have global views of the entire network. Indeed, it is not realistic to assume guaranteeing to maintain a global, synchronized view available to all nodes. In order for a variety of point centrality metrics to be more relevantly used in diverse applications, we need to think of approximation methods to take advantage of various point centrality metrics without introducing high inaccuracy due to the reduced size of a network each node to consider for the estimation of its centrality value.

IV. GRAPH CENTRALITY METRICS

In Section III, we surveyed an individual node's centrality. Now we look into the centrality of a given graph, which characterizes the centrality of all nodes in an entire network, not just points (or vertices). Hence, graph centrality can be used as an indicator to represent how nodes in a network are connected as a whole. We discuss 15 graph centrality (GC) metrics in this section.

A. MEASURES OF GRAPH CENTRALITY

1) DISTANCE-BASED GC

This measures the distances between all pairs of vertices in order to measure the *compactness* of a network. The distance-based GC is defined in [82] as:

$$C_{\text{distance-GC}}(\mathcal{G}) = \sum_{u\in\mathcal{V}}\sum_{v\in\mathcal{V}} d(u,v), \tag{43}$$

Shimbel [7] earlier developed this metric, calling it *dispersion* and interpreting it as measuring a vertex's *accessibility* to the graph $\mathcal{G}$. The average shortest path [99] is a similar metric in order to compare the breadth of a network at different scales.

2) DEGREE-BASED GC

This metric measures the relative dominance of a single vertex in a network. Nieminen [158] defined this metric as:

$$C_{\text{deg-GC}}(\mathcal{G}) = \sum_{v\in\mathcal{V}} \binom{1 + d^* - d(v)}{2}, \tag{44}$$

where d^* denotes the maximum degree in the graph $\mathcal{G}$. The maximum sum of the differences between the largest centrality and all other centralities can be derived as follows: The maximum possible degree in a graph with n nodes is $n-1$. So this metric can be normalized by scaling by the maximum degree-based GC, or $(n-1)\binom{n-1}{2}$. Freeman [39] proposed an alternative normalization based on differences of the degrees (as opposed to the binomial of differences), expressed as $C_{\text{norm-deg-GC}}(\mathcal{G}) = \frac{\sum_{v\in\mathcal{V}}[d(v^*)-d(v)]}{n^2-3n+2}$.

TABLE 3. Meanings, computations, and complexities of global point centrality metrics.

Centrality name	Meaning	Eq. No.	Complexity	Ref. No.
Improved method	An improved version of k-shell by ranking nodes with the same k-shell and using a Binary Search Tree to find the distance between two nodes	(25)	$O(n^2 \log n)$	[132]
Betweenness	Influence of a node as a broker using Floyd-Warshall algorithm using Johnson's algorithm or Brandes' algorithm with a weighted graph using Johnson's algorithm or Brandes' algorithm with a unweighted graph	(26)	 $O(n^3)$ $O(n^2 \log n + mn)$ $O(mn)$	[39]
L-betweenness	An improved version of betweenness centrality by only considering the pair whose distance smaller than L	(27)	sublinearly in $O(mn)$	[136]
Flow betweenness	An amount of network flow going through a node	(28)	$O(m^2 n)$	[60], [138]
Random-walk betweenness	Influence of a node based on the transmit speed of flow to a node with random walk	(29)	$O((m+n)n^2)$	[138]
Routing betweenness	Influence of a node based on the expected number of packet passing through a node	(31)	$O(n^2 m)$	[141]
Load	When all nodes send a packet to every other node along with a shortest path, the number of packets passing through a node; used Dijkstra algorithm	(30)	$O(mn)$	[139], [141]
Closeness	A reciprocal of the sum of distances between a node and all other nodes	(32)	$O(mn)$	[156], [157]
Information	A node's importance based on all possible paths going through the node	(33)	$O(n^3)$	[26]
Current-flow betweenness & closeness	A node's importance based on the amount of information spread over a network as an electric current	(34)-(35)	$O(n^3)$ for $m < n^2$	[142]
Residual closeness	An alternative version of closeness metric to measure the vulnerability in the graph (used Floyd-Warshall algorithm for a path distance)	(36)	$O(n^3)$	[143]
Spatial	A node's importance in terms of how efficiently it provides a route between two nodes going through the node (used Breadth First Search for finding a path between two nodes)	(37)	$O(n(n+m))$	[147]
AHP-based	A node's influence by ranking nodes based on multiple centrality metrics, such as degree, betweenness, or closeness	(38)	$O(n^3)$	[148]
Generalized degree and shortest paths	A node's importance by combining degree, closeness and betweenness metrics with a corresponding weight given	(39)	$O(n^2)$ for degree, $O(n^3)$ for closeness and betweenness	[152]
Weight neighborhood	A node's importance in diffusion using a benchmark centrality metric (ϕ), such as degree, betweenness, or k-shell	(40)	$O(nm)$	[153]
Percolation	A node's vulnerability to being percolated	(41)	$O(n^3)$	[154]
Eccentricity	A node's influence measured based on the maximum distance between the node and other nodes (i.e., lower eccentricity refers to high influence); used Floyd-Warshall algorithm for finding a path between two nodes	(42)	$O(n^3)$	[155]

(Notations: n is the total number of nodes and m is the number of edges.)

3) BETWEENNESS-BASED GC

This metric is calculated by the mean difference between the maximum betweenness and all other betweennesses [39], expressed as:

$$C_{\text{bet-GC}}(\mathcal{G}) = \frac{\sum_{v\in\mathcal{V}}[C'_{\text{bet}}(v^*) - C'_{\text{bet}}(v)]}{n-1} = \frac{\sum_{v\in\mathcal{V}}[C_{\text{bet}}(v^*) - C_{\text{bet}}(v)]}{n^3 - 4n^2 + 5n - 2}, \quad (45)$$

where v^* is the node with maximum betweenness and $C'_{\text{bet}}(\cdot)$ is the normalized betweenness [82].

4) FLOW BETWEENNESS-BASED GC

This metric determines the centrality of a weighted (or valued) graph based on the difference between the greatest maximum flow in the network and the maximum flow of all other nodes. This is computed by [60] as:

$$C_{\text{flow-bet-GC}}(\mathcal{G}) = \frac{\sum_{v\in\mathcal{V}}[C'_{\text{flow-bet}}(v^*) - C'_{\text{flow-bet}}(v)]}{n-1}, \quad (46)$$

where v^* is the vertex with maximum flow betweenness and $C'_{\text{flow-bet}}(\cdot)$ refers to the normalized flow centrality based on Eq. (28).

5) CLOSENESS-BASED GC

Freeman [82] generalized the closeness-based graph centrality measure based on the previous trials [4], [156]. This metric can be simply derived based on the normalized closeness metric, $(n-1)C_{\text{closeness}}(v)$, from Eq. (32) by:

$$C_{\text{close-GC}}(\mathcal{G}) = \frac{\sum_{v\in\mathcal{V}}[C'_{\text{closeness}}(v^*) - C'_{\text{closeness}}(v)]}{\max\sum_{v\in\mathcal{V}}[C'_{\text{closeness}}(v^*) - C'_{\text{closeness}}(v)]} = \frac{\sum_{v\in\mathcal{V}}[C'_{\text{closeness}}(v^*) - C'_{\text{closeness}}(v)]}{(n^2 - 3n + 2)/(2n-3)}, \quad (47)$$

where v^* is the vertex with the largest closeness centrality.

6) RECIPROCITY

Newman *et al.* [13], [14] defined *network reciprocity* based on the number of bidirectional edges between two nodes over the total number of edges in a network. In directed networks,

for an edge from node i to node j, if there is an edge from node j to node i, it is said the edge from node i to node j is reciprocated, which is also called *co-links* in the World Wide Web context [107]. Formally put, the reciprocity can be denoted by:

$$C_{\text{reciprocity}} = \frac{\sum_{ij} A_{ij}A_{ji}}{m} = \frac{\text{Tr}(\mathbf{A}^2)}{m}, \tag{48}$$

where the number of edges m, in this case, refers to the sum of the number of bidirectional edges and the number of unidirectional edges.

7) k-COMPONENT

This metric refers to a maximal subset of nodes where each node can reach from each of other nodes based on minimum k paths that are vertex-independent. Note that two paths are said to be *vertex-independent* if they do not share any of the same vertices [13]. A variant of the k-component can be identified based on edge-independent paths, implying that removing less than k edges cannot make the component disconnected [13].

8) k-CLIQUE

A *clique* refers to a maximum subset consisting of vertices in an undirected network where each member of the subset is directly connected to each other [159], [160]. If the size of the clique is large, it represents a highly *cohesive* network with close connectedness between each other [13].

9) k-PLEX

This metric relaxes the condition of the clique as we cannot find a perfect clique in reality. A k-plex refers to the maximum size of the subset of n vertices in a network where each vertex is connected with at least $n - k$ other vertices [159]. The 1-plex is indeed a clique.

10) k-CORE

This metric is a very close concept to the k-plex. It refers to the maximum size of a subset consisting of vertices that have minimum k connections with other vertices in the subset. In this sense, the k-core is a $(n - k)$-plex. But given a k value, the set of all k-cores is not the same as that of all k-plexes because n is different for a different k-core. Further, different from k-plexes, each k-core is distinct because when two k-cores share one or more vertices, a single, larger-sized k-core can be formed [13], [159].

11) AVERAGE CLUSTERING COEFFICIENT

Based on the mean of (local) clustering coefficient for a given graph, Watts and Strogatz [99] also defined the *average clustering coefficient* (ACC) as:

$$ACC(\mathcal{G}) = \frac{\sum_{v\in\mathcal{V}} C_{\text{clustering}}(v)}{n}, \tag{49}$$

where $C_{\text{clustering}}(v)$ is the local clustering coefficient of node v [99]. An alternative *global clustering coefficient* based on connected triplets of vertices determines the transitivity of the network [13], [161]. This metric is based on the ratio of the 3-cycles in the graph to the number of connected triples and is defined as $GCC(\mathcal{G}) = \frac{\text{Tr}(\mathbf{A}^3)}{\sum_{v\in\mathcal{V}} d(v)(d(v)-1)}$.

12) DEGREE ASSORTATIVITY

Newman [162] first defined the *assortativity* of a network as a graph measure to represent to what extent nodes are associated with other nodes in terms of network structural characteristics, such as degree, betweenness, node weight, node coreness as well as node characteristics, according to some context such as ethnic, language, and/or culture. Degree assortativity can be simply defined based on the linear correlation coefficient between two nodes' excess degrees where a node's excess degree is its degree minus 1 (i.e., $d(v)-1$), also known as the remaining degree of the node and the excess degrees are random variables. The degree assortativity, ρ_D, is given by:

$$\rho_D = \frac{\sum_{jk} jk(e_{jk} - q_j q_k)}{\sigma_q^2}, \tag{50}$$

where e_{jk} refers to the joint excess degree probability for nodes with excess degrees j and k. q_k is a normalized distribution of a randomly selected node, given by $q_k = \frac{(k+1)p_k}{\sum_j jp_j}$, where σ_q is the standard deviation of the distribution q_k. Newman extended the definition of degree assortativity in unweighted, directed networks, as $\rho_D = \frac{\sum_{jk} jk(e_{jk} - q_j^{in} q_k^{out})}{\sigma_{in}\sigma_{out}}$, where e_{jk} indicates the probability that a node with out-degree k and a node with in-degree j is connected for $k, j \in \mathbb{N}$, $q_j^{in} = \frac{(j+1)p_{j+1}^{in}}{\sum_j jp_j^{in}} = \frac{(j+1)Pr[D_{in}=j+1]}{E[D_{in}]}$ is the normalized excess in-degree distribution where D_{in} is the in-degree for a randomly selected node, q_k^{out} is defined similarly, and σ_{in} and σ_{out} are the standard deviations of q_j^{in} and q_k^{out}, respectively. Noldus and Van Mieghem [163] discussed multi-layered assortativity to be applied in directed networks, including: (1) *in-degree assortivity* measuring the tendency of a particular in-degree node that is connected to the same in-degree or different in-degree nodes; (2) *out-degree assortativity* estimating the trend of a particular out-degree node's connectedness with the same out-degree or different out-degree nodes; and (3) *overall assortativity* calculated based on both in-degree assortativity and out-degree assortativity.

13) LOCAL ASSORTATIVITY

Piraveenan *et al.* [165] defined *local assortativity* to measure an individual node's assortativity based on its degree and its neighbors' degree. The local assortativity is measured by:

$$\rho(v) = \frac{(j+1)(j\bar{k} - \mu_q^2)}{2m\sigma_q^2}, \tag{51}$$

where j is the excess degree of node v (i.e., $d(v) - 1$), $\bar{k}$ is the average excess degree of node v's neighbors (i.e., $[\sum_{u\in N(v)}(d(u) - 1)]/d(v)$), σ_q is the standard deviation of the distribution of excess degree j over all nodes in the

TABLE 4. Meanings, computations, and complexities of graph centrality metrics.

Centrality name	Meaning	Complexity	Eq. No.	Ref. No.
Distance-based GC	Sum of distances between each vertex and all other vertices using the Breadth First Search	$O(n(n+m))$	(43)	[82], [7]
Degree-based GC	Maximum sum of differences between the largest centrality and all other centralities	$O(n^2)$	(44)	[158]
Betweeness-based GC	Mean difference between the maximum betweenness and all other nodes' betweenness; used Floyd-Warshall algorithm	$O(n^3)$	(45)	[39], [82]
Flow betweenness-based GC	Difference between the highest maximum flow with the highest betweenness and maximum flow of all other nodes	$O(n^4)$	(46)	[60]
Closeness-based GC	Mean difference between the maximum closeness metric and all other closeness	$O(n^3)$	(47)	[82]
Reciprocity	Number of bidirectional edges between two nodes over the total number of possible edges	$O(n^2)$	(48)	[14]
k-component	A maximal subset of nodes where each node can reach the other nodes in the subset based on minimum k paths that are vertex independent	$O(Fn^2)$	-	[13]
k-clique	A maximal subset of vertices where each vertex of the subset is directly connected to each other	$O(n^3)$	-	[159], [160], [13]
k-plex	A maximal subset of n vertices where each vertex is connected to minimum $n-k$ other vertices	$O(n^3)$	-	[159]
k-core	A maximum size of the subset where each vertex is connected to minimum k other vertices	$O(n+m)$	-	[13], [159]
Average clustering coefficient	Mean of a local clustering coefficient of a graph	$O(nd_{max}^2)$	(49)	[99], [161], [13]
Degree assortativity	Linear correlation coefficient between two nodes' excess degrees	$O(n^2)$	(50)	[162], [164], [163]
Local assortativity	An individual node's assortativity based on the node's degree and its neighbor's degree	$O(n^2)$	(51)	[165]
Graph curvature	Negative curvature of the graph as a whole to identify congestion	$O(n^2(n+m))$	(52)	[105], [166], [167], [168]

(Notations: Given a given network G, n is the total number of nodes, m is the number of edges, p is the propagation probability of influence, l is the number of iterations, and F is the time complexity to find the maximum flow between two vertices in a graph G.)

network, and μ_q is the average excess degree. Note that the sum of all local assortativities is the network assortativity, $\rho = \sum_{v\in\mathcal{V}} \rho(v)$.

14) GRAPH CURVATURE

One hypothesis to explain the phenomenon observed in many large networks of traffic congestion occurring at a core set of nodes in the network is that the network as a whole is negatively curved. Evidence supporting this hypothesis includes the ease in embedding networks in hyperbolic space or deriving various properties using hyperbolic network models [105]. If the network is negatively curved, then routing paths influenced by shortest path selection are somewhat forced to traverse this core, leading to congestion. Point centralities are useful in potentially identifying this core set, but they do not measure the network curvature of the graph as a whole. To address this problem, Narayan and Saniee [166] developed a large-scale curvature measure by adapting to graphs the "δ-thin triangle condition" [167] that defines negative curvature. For any triple of nodes i, j, k, we define the distance function from any other node h to the triangle of nodes by $D(h; i, j, k) = \max\{d(h; i, j), d(h; i, k), d(h; j, k)\}$ where $d(h; u, v)$ is the minimum distance from the node h to the geodesic between u and v. Then, the curvature of a network with respect to the triple, $\delta_{i,j,k}$, can be defined as:

$$\delta_{i,j,k} = \min_h D(h; i, j, k). \tag{52}$$

An infinite network is negatively curved (hyperbolic), if $\delta = \max_{i,j,k} \delta_{i,j,k} < \infty$. Obviously, finite networks would not satisfy this condition, hence comparing δ to the perimeter length of the triangle formed from the geodesics among the triple (i, j, k). This ratio does not exceed $3/2$ for constant non-positively curved Riemannian manifolds [168]. To relax the constraint that every triple satisfies this condition and for computational reasons, Narayan and Saniee [166] considered a random sampling of triples and determine if the ratio δ_Δ/ℓ converges for large $\ell = \min\{d(i, j), d(i, k), d(j, k)\}$.

15) ASYMPTOTIC COMPLEXITY OF GRAPH CENTRALITY METRICS

In Table 4, we summarized the meanings, equations, and asymptotic complexity in big-O of the graph centrality metrics surveyed in our paper. We can observe the majority of GC metrics have their asymptotic complexity from $O(n^2)$ to $O(n^4)$. Most of the GC metrics are measured based on the distances between two nodes or the difference between certain centrality values of two nodes. Therefore, GC metrics represent how a network is clustered based on certain criteria to represent the relationships between nodes in the network.

B. DISCUSSIONS: GRAPH CENTRALITY METRICS

Recall that a graph centrality (GC) can represent how nodes in a network are connected to each other with a single value metric. Due to this merit, we can use GC metrics to indicate

a certain level of system performance or conditions if we can identify some relationships between a given GC and system conditions. For example, if we can observe higher vulnerability to epidemic attacks in a network with higher betweenness-based GC, then we can use the GC as a metric to indicate the extent of network vulnerability to the network. We further discussed this through our experiments in Section VII. The experimental results in Tables 8 and 9 (see Section VII) showed that higher network vulnerability due to either non-infectious or infections attacks is observed when a network shows lower GC because high network connectivity introduces network vulnerability due to the nature of interconnectivity, resulting in cascading failures, as discussed in [12].

V. GROUP CENTRALITY METRICS

When a group of nodes is selected for many of the problems in the application space (e.g., influence maximization, network destruction), simply selecting the top-K ranked nodes is a naïve approach. Many networks exhibit assortativity, with respect to degree or another centrality, or redundant clustering. A simple example demonstrating the problem with top-K selection strategy is the observation of the importance of the k-shell (certainly for influence maximization), as the top-K nodes all may reside in the same k-shell and be neighbors. k-shell based centrality approaches would only push the selected nodes to the edge of the top k-shell, which may be highly localized instead of distributed throughout the network.

One approach to resolving this issue to iteratively select a single node and recalculate the centrality measure for the remaining network excluding the selected node(s). This strategy has been studied for network robustness [15] and the recalculation can be trivial for certain measures (e.g., degree, coreness). For other measures, this recalculation may be expensive. Hence less costly approaches have been developed, seeking to discover a more close-to-optimal set of k nodes.

A. MEASURES OF GROUP CENTRALITY

1) DEGREEDISTANCE

Sheikhahmadi *et al.* [169] introduced a degree-distance metric to ensure the selected nodes are well-dispersed in the network. The strategy first computes the degree of each node and selects the node with the highest degree. It then excludes the selection of all nodes within a chosen threshold distance t_{td} from any of the previously selected nodes and selects the node with the highest degree. Given a current set of selected seed nodes S (e.g., selected at random), the next selected node is chosen to be

$$v = \underset{u\in\mathcal{V}|d(u,w)\geq t_{td}\forall w\in S}{\operatorname{argmax}} d(u). \tag{53}$$

Since this threshold distance can omit from the potential selection of high degree nodes that are within the threshold distance but have limited common neighbors (or neighbors of neighbors) with the previously selected nodes, the authors introduced two improvements to *DegreeDistance*. The first improvement of DegreeDistance (FIDD) does not exclude a node v within the threshold distance, provided the number of common neighbors and common neighbors of neighbors with previously selected nodes in S is below a chosen threshold θ. The second improvement of DegreeDistance (SIDD) adds another check to determine an influence score $\mathbb{P}(u,v) + \sum_{w\in N(u)\cap N(v)}(\mathbb{P}(u,w)\cdot\mathbb{P}(w,v))$, where $\mathbb{P}(u,v)$ is the activation probability that u will influence v. Nodes within the threshold distance with influence above some threshold β are excluded from being selected for inclusion in S even when the common neighbors are below the threshold θ. Essentially, sufficient pathways exist for the node to be affected by a seed node indirectly.

2) SINGLEDISCOUNT

This is essentially the iterative recalculation of degree. Chen *et al.* [170] used this basic heuristic to compare against several greedy approaches to estimate the cascade models of Kempe *et al.* [171]. The node with maximum degree is selected for the seed set S (ties broken randomly). Each neighbor of a selected node has a unit value reduction in its degree. This selection can be represented by:

$$v = \underset{u\in\mathcal{V}\setminus S}{\operatorname{argmax}} d(u) - |N(u)\cup S|, \tag{54}$$

where $d(u) - |N(u)\cup S| = |N(u)| - |N(u)\cup S|$ is the degree of node u excluding the current links to the seed set S.

3) DEGREEDISCOUNT

The SingleDiscount approach ignores the probability that a node may be affected by a neighbor in the seed set. Chen *et al.* [170] constructed an alternate heuristic to account for this and better match the *independent cascade model* of Kempe *et al.* [171]. Given a small propagation probability of p, we assume that t_v neighbors of v are already in the seed set, and that $d(v) = O(1/p)$ and $t_v - O(1/p)$. Then, the expected number of additional vertices in $N(v)$ that will be influenced by the selection of v can be shown to be $1 + (d(v) - 2\,t_v - (d(v) - t_v)t_v p + o(t_v))\cdot p$. This is derived via the probability $(1-p)^{t_v}$ that v would not be influenced by nodes already in the seed set and the expected number of vertices $1 + (d(v) - t_v)\cdot p$ that v influences its neighbors that are not in the seed set. This ignores indirect influences, which would be expected to be minimal for small p. Hence, the selection criteria, using an appropriate *DegreeDiscount* is

$$v = \underset{v\in\mathcal{V}\setminus S}{\operatorname{argmax}} d(v) - 2t_v - (d(v) - t_v)t_v\cdot p, \tag{55}$$

where S is the current seed set.

4) DEGREEPUNISHMENT

To account for indirect influence from nodes in the seed set, Wang *et al.* [172] introduced a strategy that punishes nodes near the seed set. The punishment is determined by

TABLE 5. Meanings, computations, and complexities of group centrality metrics.

Centrality name	Meaning	Complexity	Metric Eq. No.	Ref. No.
DegreeDistance	Selecting nodes that have a certain distance of separation unless the number of common neighbors is limited and the influence probability is low	$O(n(n+m))$	(53)	[169]
SingleDiscount	Selecting nodes based on the maximum degree minus the number of shared neighboring nodes among the nodes in a seed set	$O(n^2)$	(54)	[170]
DegreeDiscount	Selecting nodes based on their degrees reduced by the influence probability of neighboring nodes in a seed set	$O(p \log n + m)$	(55)	[170]
DegreePunishment	Selecting nodes based on their degrees minus the number of shortest paths they are on in a seed set.	$O(l(n + \langle k \rangle^2))$	(56)	[172]
Collective Influence	Selecting nodes based on the hierarchical corona of hubs, similar to the volume centrality	$O(n \log n)$	(57)	[173], [174]

(Notations: Given a given network G, n is the total number of nodes, m is the number of edges, $\langle k \rangle$ is the mean degree of nodes, l is the number of iterations, and p is the propagation probability of influence. Note that a seed set is selected at random or based on a certain conditions, such as the extent of a certain point centrality, among the nodes in a given network G.)

how many shortest paths the node is on, the penalty is more severe if the node is closer to a seed and, consequently, closer to the seed on the paths. This punishment is $p_{u\to v} = d(u)\sum_{h=1}^{r-1}(\mathbf{A}^h)_{uv}\omega^h$, where ω is a weaken factor (typically assigned to be the propagation probability) and r is the radius of influence or length of the considered paths. Then given the current seed set S, the *DegreePunishment* selection of the next node is given by:

$$v = \underset{v\in\mathcal{V}\setminus S}{\operatorname{argmax}}\, d(v) - \sum_{u\in S} p_{u\to v}. \tag{56}$$

The complexity of this process grows with the radius r of the paths from the seed set, so Wang *et al.* [172] limited the radius to $r = 2$ in their simulations.

5) COLLECTIVE INFLUENCE

Morone and Makse [173] introduced a scheme to capture the *collective influence* (CI) of a set of nodes using the concept of optimal percolation. The influence of a single node is determined by its corona, defined in a similar manner as volume centrality (see Eq. (4)). This influence of a node v is $C_{\text{collective-inf}}(v,\ell) = (C_{\deg}(v) - 1)\sum_{u\in B(v,\ell)}(C_{\deg}(u) - 1)$, where $B(v,\ell)$ is the set of nodes within the distance of ℓ from node v. Hence, given the current seed set S, the next node selected is

$$v = \underset{v\in\mathcal{V}'=\mathcal{V}\setminus S}{\operatorname{argmax}}\, C_{\text{collective-inf}}(v,\ell), \tag{57}$$

where the collective influence is measured based on the remaining graph consisting of nodes where nodes in S are removed. Morone *et al.* [174] also provided a stopping criteria for their approach by updating an estimate of a lower bound on the minimum eigenvalue of the non-backtracking matrix when a fraction of q nodes are removed. This estimate is given by $\lambda(\ell;q) = \left(\frac{\sum_v C_{\text{collective-inf}}(v,\ell)}{n\langle k\rangle}\right)^{1/(\ell+1)}$, where $\langle k\rangle$ is the mean degree of the original network. When $\lambda(\ell;q) = 1$, the selection process is finished.

6) ASYMPTOTIC COMPLEXITY OF GROUP CENTRALITY METRICS

In Table 5, we summarized the meanings, equations, and asymptotic complexity in the big-O notation of the group centrality metrics surveyed in this section. When the range of the number of edges m is considered as $n-1 \le m \le n^2$, the complexity can range from $O(n)$ to $O(n^3)$. Unsurprisingly, the base centrality for any group centrality will be less than that for the group centrality. The literature in this area currently focuses on selecting a set of optimal nodes based on the degree and tries to limit the case where neighboring nodes are selected. Any of these approaches can be extendable to other local point centrality metrics in a similar manner.

B. DISCUSSIONS: GROUP CENTRALITY METRICS

The core idea of group centrality metrics is how to select a set of nodes with high influence without making them localized in a certain region. That is, the group centrality metrics want to select a set of nodes with high influence that are distributed in a given network by excluding the overlapped influence between nodes. To do this, group centrality metrics select a set of nodes as the seed nodes (which is often selected based on the extent of its centrality, such as a node's degree) and use them to estimate the penalty based on the distance between the node and a seed (i.e., a closer distance leads to a higher penalty). This implies that how to select a set of seed nodes is critical in selecting a set of nodes as a group. Group centrality metrics are more efficient than point centrality metrics when an application is to select a set of nodes based on the importance in a given network topology.

VI. EVOLUTION OF CENTRALITY METRICS

To observe the overall trends of how centrality metrics have evolved over time, we showed the number of publications studying the three types of centrality metrics (i.e., point, graph, and group selection centrality metrics) from the 1960s or earlier until the 2010s in Fig. 2. From the figure, we can clearly notice that various types of centrality metrics have been significantly studied since the 2000s and more actively

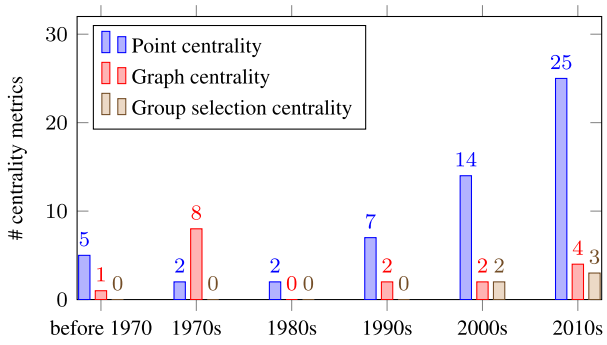


FIGURE 2. Centrality metrics developed under each category from the most recent five decades and before.

in the 2010s. More noticeably, most studies are more interested in developing point centrality metrics. Although we can observe the dominant trends of studying point centrality metrics across all times, in the 1970s, it is interesting to observe that more graph centrality metrics are developed than other types of centrality metrics.

For the easy reference of each metric discussed in Sections III-V, we summarized the list in Table 6 based on publication year in order to capture the evolution of centrality metrics. The most common centrality metrics used in the literature (e.g., degree, betweenness, closeness, eigenvector centrality) were proposed before the 1990s and a large volume of new point centrality metrics have been studied in the 2000s and 2010s. To take advantage of those various types of centrality metrics that are recently proposed, we hope that more studies consider various centrality metrics proposed in the literature through our survey paper.

VII. NETWORK RESILIENCE ANALYSIS OF THE SURVEYED CENTRALITY METRICS

A. EXPERIMENTAL SETUP

This section explains the experimental setup used for evaluating the performance of each centrality metric surveyed in this work in terms of the size of the giant component as the indicator of network resilience. To be specific, we provide datasets, metrics, and attack scenarios used for evaluating the surveyed centrality metrics in this work.

1) DATASETS

We selected the following real datasets for network topologies used in the performance demonstration of the surveyed centrality metrics:

- *Directed Network Topologies:* (1) The UCI Social Network[2] [178] is a collection of interactions from private messages sent over an online social network at The University of California, Irvine. (2) The Rocketfuel Network[3] [179] is a snapshot of router connections on an Internet Service Provider (ISP) topology from measurements.
- *Undirected Network Topologies:* (1) The URV Email Network[4] [180] captures the email communication for the *Universitat Rovira i Virgili* in Spain. (2) The EU Email Network[5] [181] captures the internal (or core) email communication for a large European research institution.

In Fig. 3, we described the topologies and degree distributions of all four datasets used in this work for readers to better understand the characteristics of the networks.

2) METRICS

We use the following metrics to evaluate the centrality metrics discussed in this work:

- *Size of the Giant Component:* This metric measures the fraction of nodes in the giant component. This metric is a common indicator of network resilience in the Network Science [12].
- *Mean Fraction of Infected Nodes:* This metric measures the mean number of infected nodes by an initial attacker.
- *Running Time:* This measures the simulation time in seconds to calculate the centrality metrics in the given datasets.

For graph centrality metrics, we surveyed 15 metrics in Section IV. Since the range of each metric varies, we cannot compare their maximum values. However, we can at least investigate whether the value of each metric increases or decreases depending on how many nodes are removed at random and accordingly the size of the giant component. In order to easily observe this, we devised a metric called the *relative graph centrality* (RGC) value, which is computed by:

$$\text{RGC} = \frac{GC - GC'}{GC}, \tag{58}$$

where GC is the value of a given graph centrality (GC) from the original network with the size of the giant component being 1 and GC' is the value of a given GC after removing a certain percentage of nodes being removed at random. If we observe the RGC value increases under a smaller size of the giant component, S_g, it implies that the GC value decreases under the smaller S_g. On the other hand, if the RGC value decreases under a small S_g, this means the GC value increases under a smaller S_g. Therefore, the RGC can be used to easily detect whether a given GC can be used as a metric to represent the state of network connectivity in the given network as a whole.

3) ATTACK SCENARIOS

We consider the two attack types as:

- *Non-Infectious Attacks:* This attack type reflects node failures without infecting the node's neighbors. The practical examples include partial physical destruction of a system [182], non-critical nodes that are not functioning due to denial-of-service (DoS) attacks [183], or a

[2]https://snap.stanford.edu/data/CollegeMsg.html
[3]http://networkrepository.com/tech-routers-rf.php
[4]http://networkrepository.com/ia-email-univ.php
[5]https://snap.stanford.edu/data/email-Eu-core-temporal.html

TABLE 6. Evolution of centrality metrics from the 1960s or earlier to the 2010s.

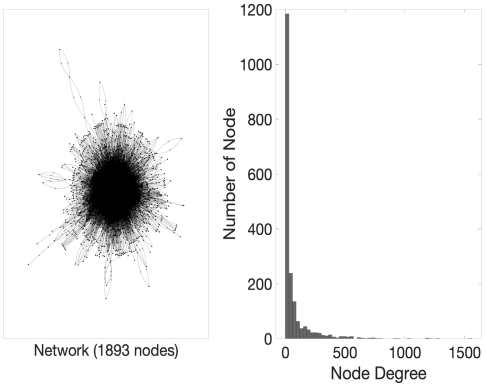
Centrality metrics	before 1970	1970s	1980s	1990s	2000s	2010s
Point centrality	Degree [1], [42], [2], [175]; Katz centrality [5]; Farness [3]; Betweenness [2], [4]; Closeness [156]	Degree [93], [82];Betweenness [82];	Eigenvector centrality [27]; Information centrality [26]	Flow betweenness [60] ; Degree [83]; Eccentricity [155]; Redundancy [98], [57]; Clustering coefficient [99]; PageRank [123]; Authority and Hub centralities [121];	Cumulative nomination [130]; SALSA [131]; Gaussian curvature on planar graphs [109]; Load centrality [139]; Curvature [107]; Degree [88], [87]; H-index [92]; Eigenvector centrality [87]; Subgraph centrality [127]; Communicability [176], [177]; Random-walk betweenness [138]; Current-flow betweenness and closeness [142]; Residual closeness [143]; straightness centrality [147]; Ricci culvature [112];	Generalized degree and shortest paths [152]; Decay centrality [145]; L-betweenness [136]; Degree [13]; Routing betweenness [141]; k-shell index or coreness [118]; Leader Rank [128]; Semi-local centrality [94]; Dynamical influence [129]; Volume [96], [97]; Gaussian curvature on planar graphs [108], [111]; Cluster Rank [102]; Diffusion centrality [126]; Mixed degree decomposition [119]; Percolation centrality [154]; Improved method [132]; Ricci culvature [113]; Neighborhood coreness [120], [118]; Contribution centrality [124]; Mapping entropy [101]; Hybrid degree [95]; Weight neighborhood centrality [153]; AHP-based centrality [148]
Graph centrality	Distance-based GC (e.g., dispersion [7])	k-clique [160]; Degree-based GC [158]; Betweenness-based GC [82]; Closeness-based GC [82]; k-clique [159]; k-plex [159]; k-core [159] ; Distance-based GC (e.g., compactness [82])		Flow betweenness-based GC [60]; Average clustering coefficient [99]	Reciprocity [14]; Degree assortativity [162]	Local Assortativity [165]; k-component [13]; Graph curvature [166];
Group Selection centrality					SingleDistance [170]; DegreeDiscount [170]	DegreeDistance [169]; Collective influence (CI) [173]; DegreePunishment [172]

TABLE 7. Characteristics of the used datasets.

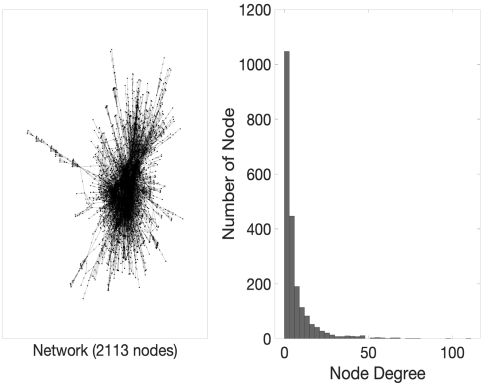
Network characteristics	UCI Social Network [178]	Rocketfuel Network [179]	URV Email Network [180]	EU Email Network [181]
Network type	Directed	Directed	Undirected	Undirected
# of nodes	1893	2113	1133	930
# of edges	59835	6632	5451	24929
Average degree	$\sim$ 63 (in+out)	$\sim$ 6 (in+out)	$\sim$ 10	$\sim$ 27
Max degree	558 (in), 1091 (out)	79 (in), 85 (out)	71	319

node accessed by an unauthorized party aiming to illegally obtain credentials [183]. The fraction of removed nodes, ϕ, is the same as the number of attackers without propagating infections.

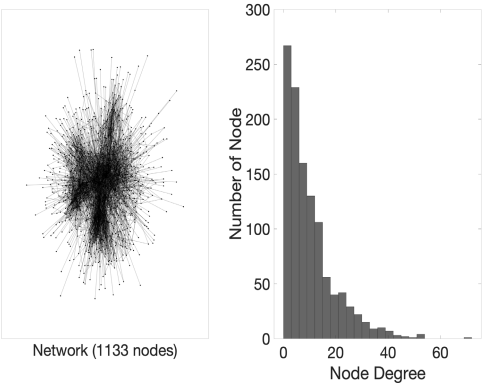
- *Infectious Attacks* Unlike the above non-infectious attack, this attack propagates infections towards other nodes. The common examples are malware or virus spreads. Botnets can propagate malware or viruses through mobile devices, which can use mobile malware such as a Trojan horse, which acts as a bot client to obtain a command and control from a remote server [183]. We model these infectious attacks by selecting the initial attackers with ϕ, a fraction of nodes being selected as initial seeding attackers. We assume that the infectious attackers follow the Susceptible-Infected-Removed (SIR) epidemic model [13]. Nodes in the susceptible state (S) refer to healthy nodes, not being infected by the attackers yet. Nodes in the infected state (I) are the compromised nodes, becoming an inside attacker, which can also replicate infections to their



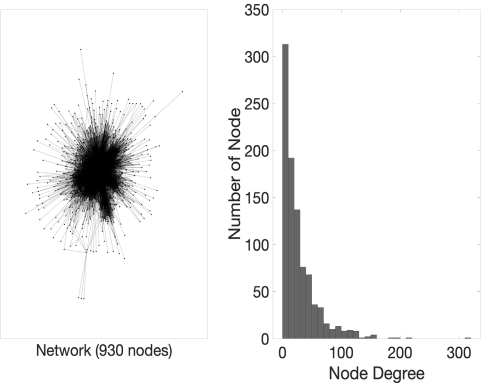
(a) UCI Social Network with 1,893 nodes and 59,835 directed edges



(b) Rocketfuel Network with 2,113 nodes and 6,632 directed edges



(c) URV Email Network with 1133 nodes and 5451 undirected edges



(d) EU Email Network with 930 nodes and 24,929 undirected edges

FIGURE 3. **Network topologies and degree distributions for the datasets used.**

neighboring nodes. Nodes in the removed state (R) are the nodes detected and isolated from the network by cutting all edges of the detected node. The compromised and detected nodes are treated as failed nodes. A susceptible node (S) can become infected (I) and later recover or be removed (R). When the size of the giant component is captured, we only consider healthy nodes, which are still in the S state. We consider the probability that a susceptible node can be infected by an infected node as the infection rate, β.

4) EXPERIMENT ENVIRONMENT SETTINGS

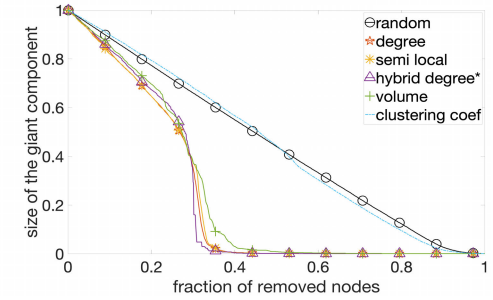
The experiment is performed using Matlab_R2019b in CentOS Linux 7.4 under x86_64. For the volume and flow betweenness centrality metrics, we used the number of hops (h) set to 2. Note that a hop in a network is defined as an edge between two connected nodes. In the group selection metrics, we used $d_{td} = 4$ in the degree distance metric and each group is defined with 10 nodes. Due to the high complexity of some metric computations (i.e., even one simulation run was expensive), we excluded the following point centrality metrics: random-walk betweenness, routing betweenness, dynamical influence, load centrality, and curvature. In the point centrality metrics, we did not show communicability centrality as it is the same as subgraph centrality when it is used to measure node centrality. In the graph centrality metrics, since reciprocity was the only metric that can be measured in a directed network, we excluded it.

B. NETWORK RESILIENCE ANALYSIS UNDER NON-INFECTIOUS ATTACKS

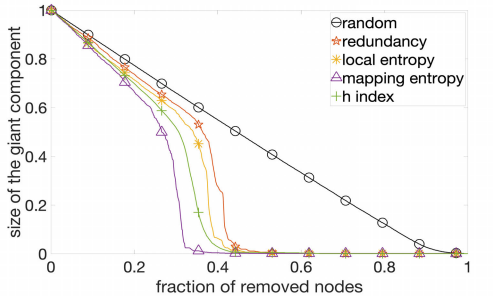
1) UNDER POINT CENTRALITY-BASED TARGETED, NON-INFECTIOUS ATTACKS

Effect of Varying the Fraction of Attackers on the Size of the Giant Component: We show the size of the giant component in the undirected URV Email Network and directed UCI Social Network in Fig. 4 and in the undirected EU Email Network and directed Rocketfuel Network in Fig. 5 when varying the fraction of removed nodes (i.e., attacked nodes), which does not infect other adjacent nodes (i.e., non-infectious attacks) selected via different point centrality metrics. This experiment shows how a targeted, non-infectious attack based on the given point centrality metric affects the size of the giant component. In Figs. 4 and 5, we observed very similar trends and obtained the following key findings:

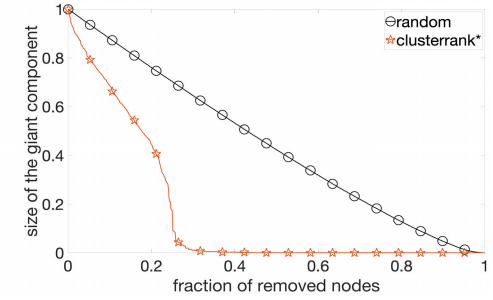
- Most targeted attacks are stronger than random attacks (notated as 'random' in black), showing a significantly lower size of the giant component.



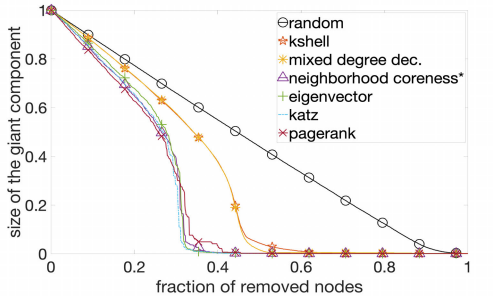
(a) *Local point centrality*: Noninfectious attacks with degree, semi-local, hybrid degree, volume, and clustering coefficient in an undirected network.



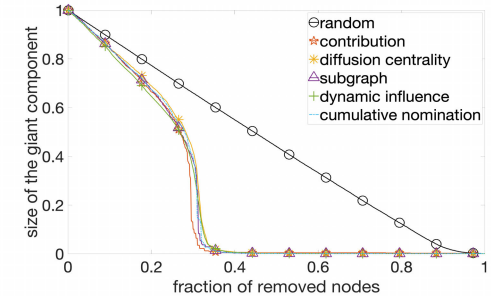
(b) *Local point centrality*: Noninfectious attacks with redundancy, local entropy, mapping entropy, and h-index in an undirected network.



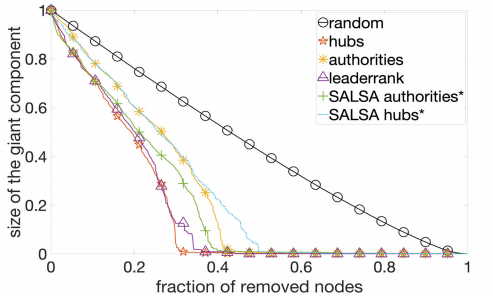
(c) *Local point centrality*: Noninfectious attacks with clusterrank in a directed network.



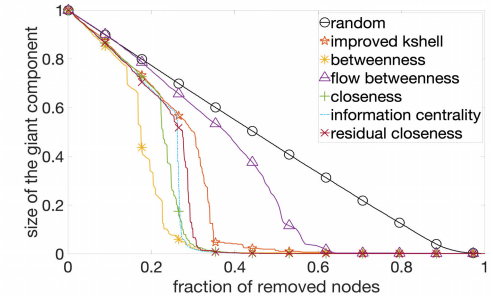
(d) *Iterative point centrality*: Noninfectious attacks with kshell, mixed degree decomposition, neighborhood coreness, eigenvector, katz, and pagerank in an undirected network.



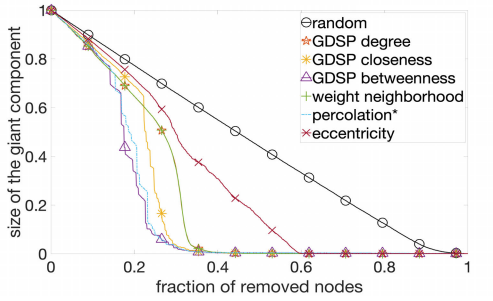
(e) *Iterative point centrality*: Noninfectious attacks with contribution diffusion centrality subgraph, dynamic influence, and cumulative nomination in an undirected network.



(f) *Iterative point centrality*: Noninfectious attacks with hubs, authorities, leaderrank, SALSA authorities/hubs in a directed network.

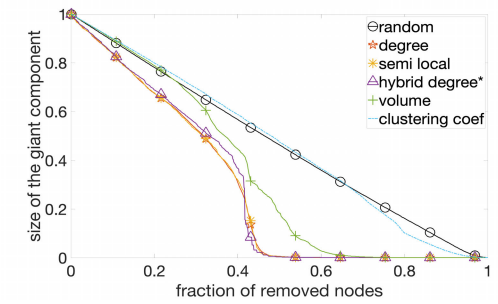


(g) *Global point centrality*: Noninfectious attacks with improved kshell, flow-betweenness, and closeness, information centrality, residual closeness in an undirected network.

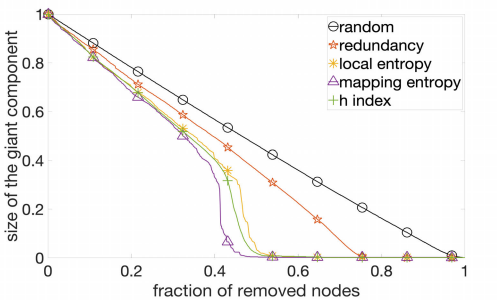


(h) *Global point centrality*: Noninfectious attacks with GDSP-degree, GDSP-closeness, GDSP-betweenness, weight neighborhood, percolation, and eccentricity in an undirected network.

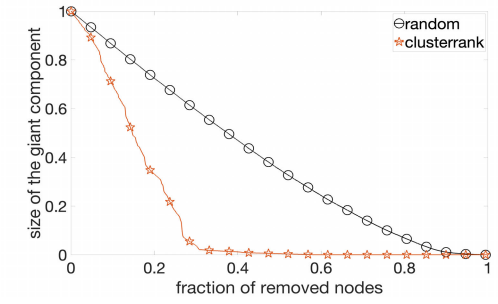
FIGURE 4. The size of the giant component after removing the initial non-infectious attacker nodes based on the point centrality metrics in the undirected URV Email Network and the directed UCI Social Network where the random node removal is included as a baseline model. (a)-(c) are for *local* point centrality metrics, (d)-(f) are for *iterative* point centrality metrica, and (g)-(h) are for *global* point centrality metrics. The star notation(*) in the legend indicates the result was obtained with only a single simulation run due to too slow running time. Otherwise, 100 simulation runs are used to obtain the mean size of the giant component.



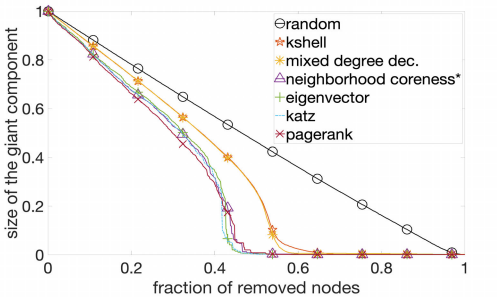
(a) *Local point centrality*: Noninfectious attacks with degree, semi-local, hybrid degree, volume, and clustering coefficient in an undirected network.



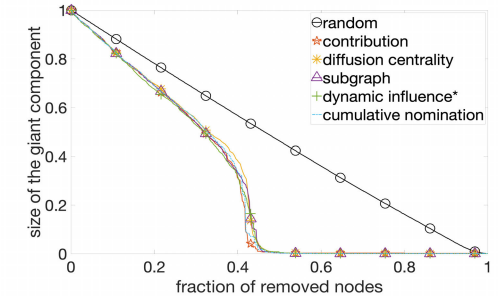
(b) *Local point centrality*: Noninfectious attacks with redundancy, local entropy, mapping entropy, and h-index in an undirected network.



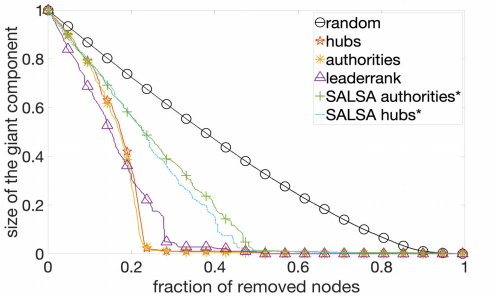
(c) *Local point centrality*: Noninfectious attacks with clusterrank in a directed network.



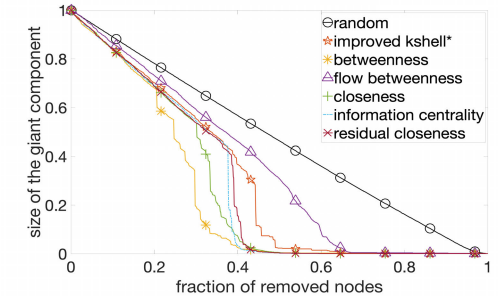
(d) *Iterative point centrality*: Noninfectious attacks with kshell, mixed degree decomposition, neighborhood coreness, eigenvector, katz, and pagerank in an undirected network.



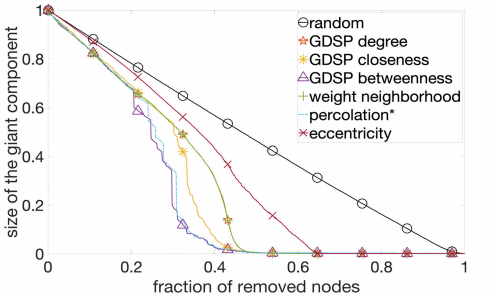
(e) *Iterative point centrality*: Noninfectious attacks with contribution diffusion centrality subgraph, dynamic influence, and cumulative nomination in an undirected network.



(f) *Iterative point centrality*: Noninfectious attacks with hubs, authorities, leaderrank, SALSA authorities/hubs in a directed network.



(g) *Global point centrality*: Noninfectious attacks with improved kshell, flow-betweenness, and closeness, information centrality, residual closeness in an undirected network.



(h) Noninfectious attacks with improved kshell, flow-betweenness, and closeness, information centrality, residual closeness in an undirected network.

FIGURE 5. The size of the giant component after removing the initial non-infectious attacker nodes based on the point centrality metrics in the undirected EU Email Network and the directed Rocketfuel Social Network where the random node removal is included as a baseline model. (a)-(c) are for *local* point centrality metrics, (d)-(f) are for *iterative* point centrality metrica, and (g)-(h) are for *global* point centrality metrics. The star notation(*) in the legend indicates the result was obtained with only a single simulation run due to too slow running time. Otherwise, 100 simulation runs are used to obtain the mean size of the giant component.

- Betweenness in (g) and GDSP betweenness in (h) show the best performance (i.e., in the sense of reducing the size of the giant component) with the network dissolved after a little more than 40% of the nodes are removed.
- Although most targeted attacks with given point centrality metrics outperform a random attack, the attack with clustering coefficient in (a) performs close to the random attack without showing a higher impact in disconnecting a given network. This is because the clustering coefficient measures the number of triangle relationships among a node's adjacent nodes, removing a node with high clustering coefficient still allows neighboring nodes to remain connected.
- The impact of removing a node is lessened if the selection criteria (or centrality) has a more local, rather than a global, scope. Therefore, removing a node with high clustering coefficient does not introduce a dramatic effect in reducing the size of the giant component.

Effect of Removing the Top 50% of Nodes Based on a Given Point Centrality Metric: Fig. 6 shows the size of the giant component after the top 50 percent of the nodes, ranked based on each point centrality, are removed. Note that this attack is not infectious so an attacked node cannot compromise adjacent nodes. From Fig. 6, we observed the following trends:

- In Fig. 6 (a) that shows the size of the giant component when non-infectious, targeted attacks are performed based on *local* point centrality metrics, except redundancy centrality, we observe a larger size of the giant component under the URV Email Network than the EU Email Network where both the URV and EU Email Networks are undirected graphs. For the clusterrank which is applied under directed networks (i.e., UCI and Rocketfuel Networks), the size of the giant component is substantially small. Recall that the EU network is denser than the URV network. The volume centrality-based attacks seem less impactful under a sparser network (i.e., URV network) while the redundancy centrality-based attacks are less impactful under a denser network (i.e., EU network) showing a larger size of the giant component.
- In Fig. 6 (b) that shows the size of the giant component when non-infectious, targeted attacks are performed based on *iterative* point centrality metrics, the diffusion centrality-based attacks performed the worst showing the highest sizes of the giant components under both undirected networks (EU and URV networks). For the iterative centrality metrics applied in directed networks, they performed worse under a sparser network (i.e., Rocketfuel network) than under a denser network (i.e., UCI network).
- In Fig. 6 (c) that shows the size of the giant component when non-infectious, targeted attacks are performed based on *global* point centrality metrics, the percolation and eccentricity centrality metrics performed the worst under both the EU and URV networks. In addition, we can easily notice that some of global point centrality metrics, such as improved method (based on k-shell), closeness, residual closeness, GDSP closeness, and weight neighborhood, showed significantly higher sizes of the giant components under the URV network, which is sparser than the EU network. This would be because lack of connectivity may limit the impact of removing nodes with high centrality.
- The effect of the point centrality on the size of the giant component depends also on the network topology. For example, for the volume centrality in (a), node removals in the EU network result in a significantly larger size of the giant component than node removals in the URV network. In addition, most point centrality metrics show a larger size of the giant component under the URV network, which is sparser than the EU network.
- In the metrics evaluated under the directed networks (see cyan and yellow bars in (a) and (b) of Fig. 6), we can clearly notice poor performance of authorities, SALSA hubs, and SALSA authorities on a sparse network as the Rocketfuel Network.

TABLE 8. Relative Graph Centrality (RGC) values of 10 GC metrics under non-infectious attacks in the undirected network datasets (EU email network, URV email network).

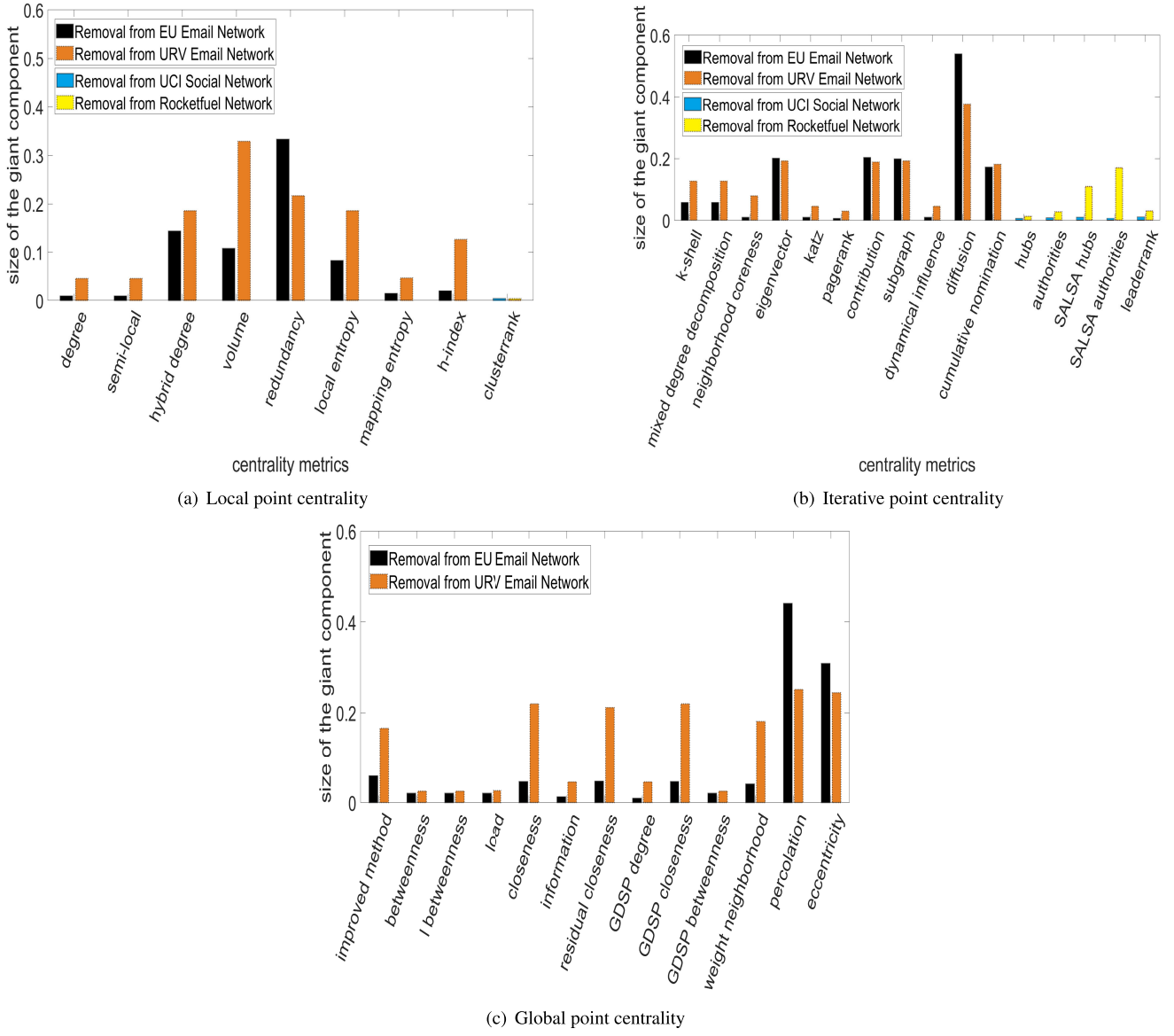
Dataset	EU Email Network		URV Email Network	
% of node removal	30%	70%	30%	70%
Size of the giant component (S_g)	∼0.7	∼0.3	∼0.7	∼0.3
distance-based	0.538	0.936	0.538	0.927
degree-based	0.396	0.756	0.423	0.830
k-component	0.109	0.477	0.105	0.34
local assortativity	0.052	0.282	0.017	0.092
graph curvature	0.028	0.152	-0.024	-0.074
average clustering	0.104	0.413	0.062	0.209
betweenness-based	-0.376	-1.427	-0.041	-0.0184
flow betweenness	-0.156	-0.603	-0.035	- 0.146
closeness-based	-0.105	-0.146	0.016	-0.005
degree assortativity	-0.015	0.052	0.057	0.101

2) UNDER GRAPH CENTRALITY-BASED TARGETED, NON-INFECTIOUS ATTACKS

For the validation of group selection centrality (GC) metrics, we considered two sets of random attacks with 30% and 70% removal of nodes in two undirected network datasets, which are the EU Email Network and URV Email Network. Since we considered random attacks in this case to investigate how the GC values are affected under two different scenarios, we observed that the size of the giant component was similar with approximately 0.3 and 0.7 for the respective cases. Since k-plex, k-clique, and k-core return a set and reciprocity needs to be applied in a directed network, we omitted the discussions of those metrics. In Table 8, we summarized the RGC values.

From Table 8, we made the following observations:

- Overall the size of giant components under different GC metrics is similar because the attacks are random.
- The effects of random attacks on the extent of GC values are different depending on each GC metric.



(a) Local point centrality

(b) Iterative point centrality

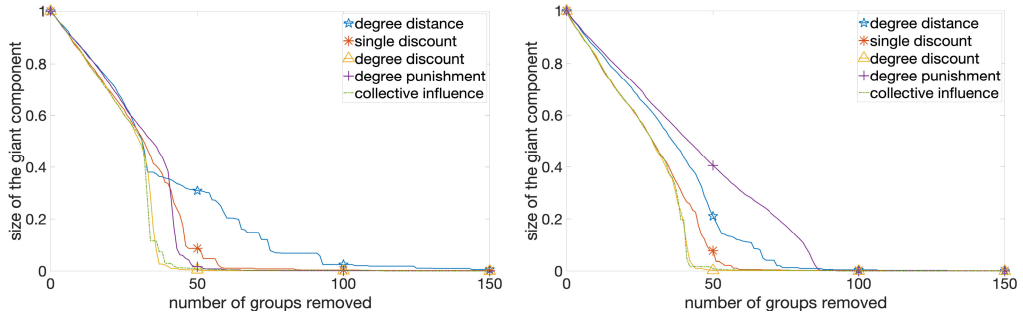
(c) Global point centrality

FIGURE 6. The size of the giant component after removing the top 50 percent of the non-infectious attackers selected based on the given point centrality metrics in both undirected networks (i.e., EU Email Network and URV Email Network) and directed networks (i.e., UCI Social Network and RocketfuelNetwork).

- We found that increasing the number of initial attackers reduces the GC value in the following graph centrality metrics: distance-based GC, degree-based GC, k-component, degree assortativity, local assortativity, and average clustering.
- We observed greater GC when increasing the number of attackers in the following GC metrics: betweenness-based GC, closeness-based GC, and graph curvature.
- The reason of showing different trends under different metrics can be explained as follows. If the GC metric measures how the node is locally connected with its close neighbors, then the GC value decreases due to the breakdown of local connections when random attacks are performed. However, if the GC metric estimates how the node is globally connected with other nodes, its value can increase as the normalization of the GC calculation depends on the size of the network. Therefore, we cannot simply rely on whether a network is dense or sparse based on the GC metric because a higher GC metric does not always necessarily imply a denser network.

3) UNDER GROUP SELECTION CENTRALITY-BASED TARGETED, NON-INFECTIOUS ATTACKS

Fig. 7 shows sizes of the giant component in both undirected networks, including the EU Email Network and URV Email Network, as the indicator of network resilience when a set of groups (where a group is defined as 10 nodes) chosen based on a given group selection metric are removed as



(a) Under noninfectious attacks in the URV Email Network (b) Under noninfectious attacks in the EU Email Network

FIGURE 7. The size of the giant component after removing a set of either non-infectious, initial attackers based a given group selection metrics in the two undirected network datasets, which are the EU Email Network and URV Email Network.

targeted, non-infectious attacks. Overall we found that attacks on denser networks (with more edges) in the EU Email Network are less severe when degree punishment is the selection criteria while attacks on larger networks (with more nodes) are less severe with degree distance.

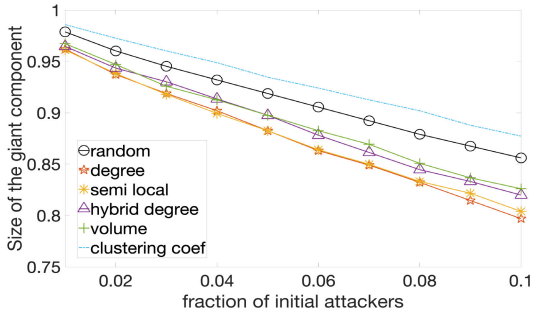
C. NETWORK RESILIENCE UNDER INFECTIOUS ATTACKS

1) UNDER POINT CENTRALITY-BASED TARGETED, INFECTIOUS ATTACKS

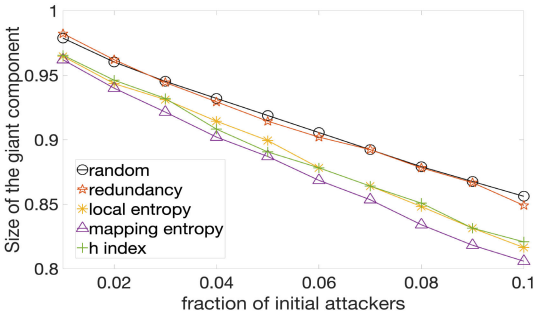
We also evaluated the performance of point centrality metrics surveyed in this work under infectious attacks. As discussed in Section VII-A3, an seeded attacker can infect neighboring nodes with an infection probability $\beta = 0.05$. Figs. 8 and 9 show the size of the giant component under targeted attacks of the URV Email Network and the UCI Social Network and the EU Email Network and Rocketfuel Network for the point centrality metrics, respectively. We varied the fraction of the initial attackers by an increment of 0.01 from 0.01 to 0.1. A node is immune to the attack if the node is attacked but is not infected based on the given infection probability, β. Note that we report results over a smaller fraction of initial attackers because of the stronger impact of infectious attacks on the size of the giant component.

From Figs. 8 and 9, we observed the following:

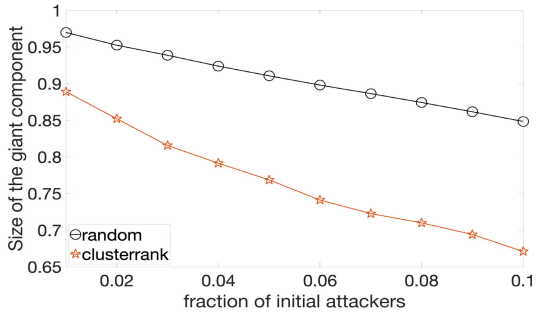
- Overall the size of the giant component linearly decreases when targeted, infections attacks are applied. Most targeted attacks reduce the size of the giant component compared to random attacks.
- Three point centrality metrics tested in this work resulted in a comparable or larger size of the giant component than random attacks. These are clustering coefficient, flow betweenness, and redundancy.
- For the clustering coefficient in (a), removing a node with high clustering coefficient has a limited effect on its local network due to high connectivity. More generally, when local neighborhoods are well connected, which is the case for nodes with high clustering coefficient, the reduction of the network is tempered.
- Similarly, since redundancy in (b) captures the overlap of a node's neighborhood with that of other nodes, the network is less likely to be dismantled because the nodes in the neighborhood remain connected.
- Note that volume centrality in (a) is estimated based on a given hop h which is set to 3 in our work. This means that even when a node with high volume centrality is removed, an infectious propagation of the attack may be limited in scope depending on the immunity of the contact or nearest neighbors.
- Lastly, the performances of betweenness in (g), pagerank in (d), and GDSP betweenness in (h) are impressive compared to other centrality metrics, resulting in a significantly smaller size of the giant component for the undirected URV Email Network. In addition, in the (directed) UCI Social Network, clusterrank in (c), leaderrank in (f), hubs in (f), and SALSA authorities in (f) are quite impressive in their performance, resulting in a significantly smaller size of the giant component, compared to other centrality metrics.
- Although the performance of Fig. 9 is very similar to that of Fig. 8, we also found some distinctive trends as follows.
  - Seeding attackers based on flow betweenness in Fig. 9 g) performs better in the EU Email Network as the fraction of initial infectious attackers increases whereas in the URV Email Network, selection based on flow betweenness performed no better than random selection, as shown in Fig. 8 (g).
  - Volume centrality-based seeding did not perform as well in the EU Email Network (Fig. 9 (a)) compared to the URV Email Network (Fig. 8 (a)). This could be because of the reason discussed earlier regarding the clustering coefficient, which also did not perform better compared to the random attack. That is, removing a node with high volume centrality may only collapse the local network of the node. This means that under dense networks, the removal of nodes with a highly connected local neighborhood does little to separate the network into smaller components.
  - We found that the network topology really affects the performance of centrality metrics. In particular, the key difference between these two datasets



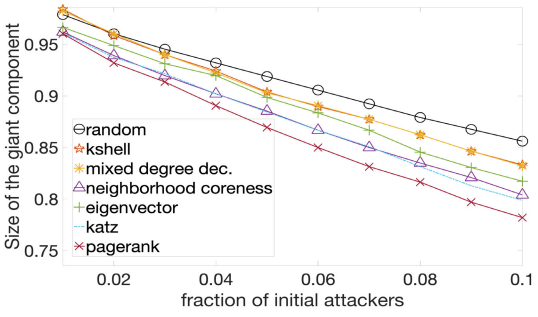
(a) *Local point centrality*: Infectious attacks with degree, semi-local, hybrid degree, volume, and clustering coefficient in an undirected network.



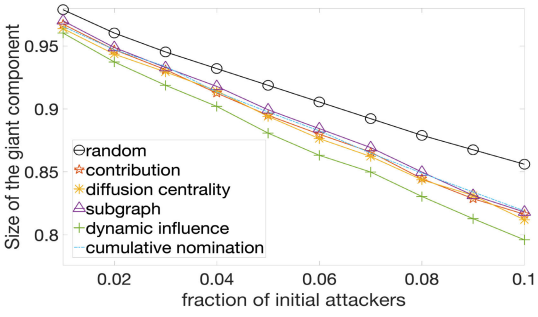
(b) *Local point centrality*: Infectious attacks with redundancy, local entropy, mapping entropy, and h-index in an undirected network.



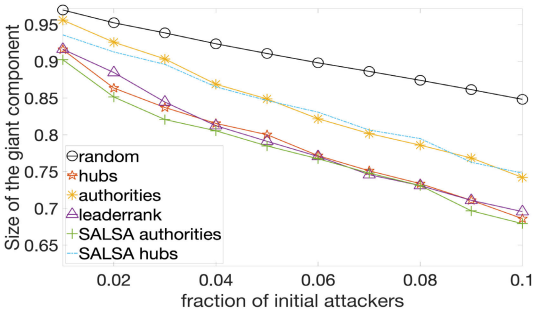
(c) *Local point centrality*: Infectious attacks with clusterrank in a directed network.



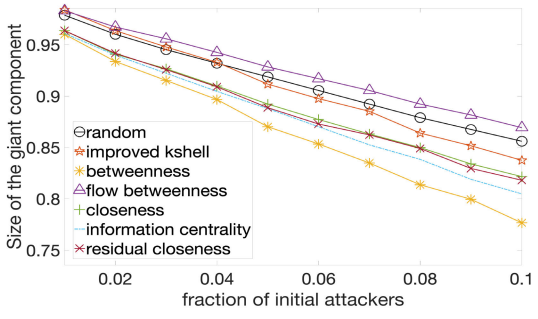
(d) *Iterative point centrality*: Infectious attacks with kshell, mixed degree decomposition, neighborhood coreness, eigenvector, katz, and pagerank in an undirected network.



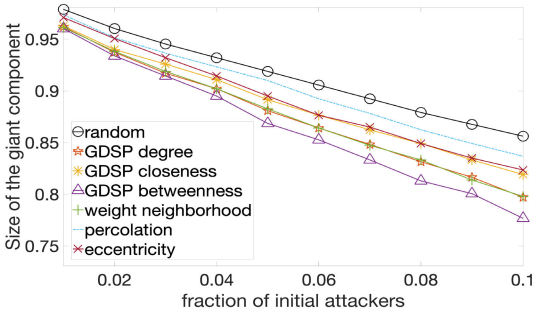
(e) *Iterative point centrality*: Infectious attacks with contribution diffusion centrality subgraph, dynamic influence, and cumulative nomination in an undirected network.



(f) *Iterative point centrality*: Infectious attacks with hubs, authorities, leaderrank, SALSA authorities/hubs in a directed network.

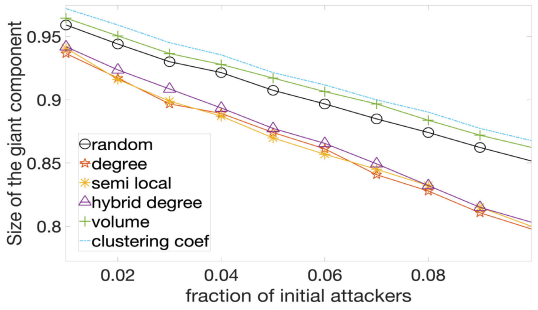


(g) *Global point centrality*: Infectious attacks with improved kshell, flow-betweenness, and closeness, information centrality, residual closeness in an undirected network.

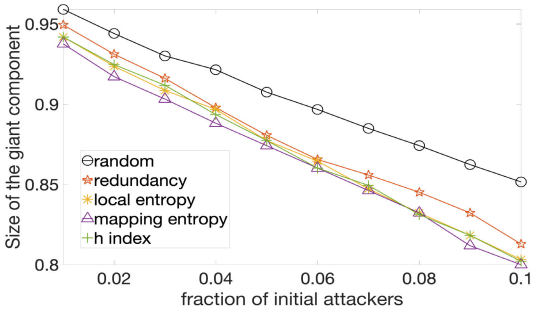


(h) *Global point centrality*: Infectious attacks with GDSP-degree, GDSP-closeness, GDSP-betweenness, weight neighborhood, percolation, and eccentricity in an undirected network.

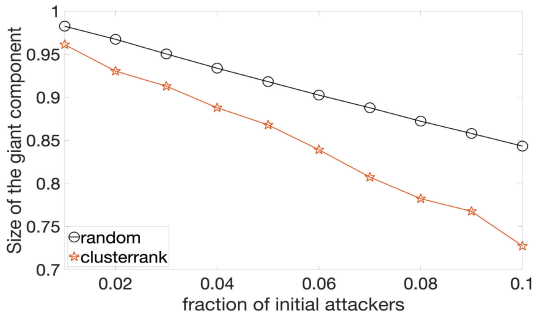
FIGURE 8. The size of the giant component after removing the initial infectious attacker nodes based on the point centrality metrics in the undirected URV Email Network and the directed UCI Social Network where the random node removal is included as a baseline model. (a)-(c) are for *local* point centrality metrics, (d)-(f) are for *iterative* point centrality metrica, and (g)-(h) are for *global* point centrality metrics. All results are based on 100 simulation runs to obtain the mean size of the giant component.



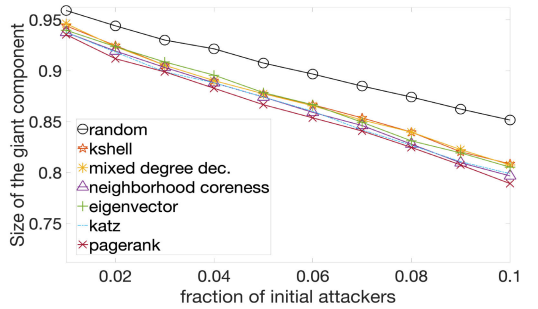
(a) *Local point centrality*: Infectious attacks with degree, semi-local, hybrid degree, volume, and clustering coefficient in an undirected network.



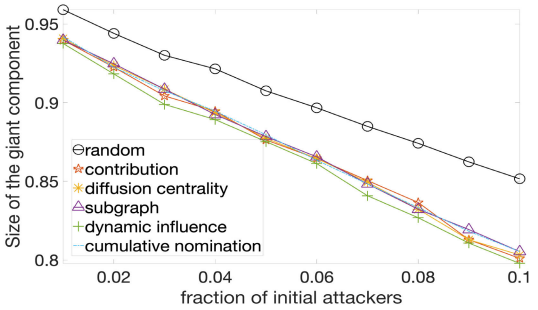
(b) *Local point centrality*: Infectious attacks with redundancy, local entropy, mapping entropy, and h-index in an undirected network.



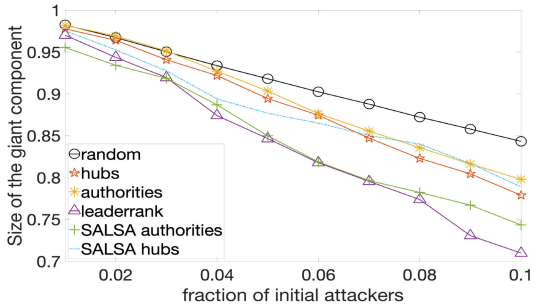
(c) *Local point centrality*: Infectious attacks with clusterrank in a directed network.



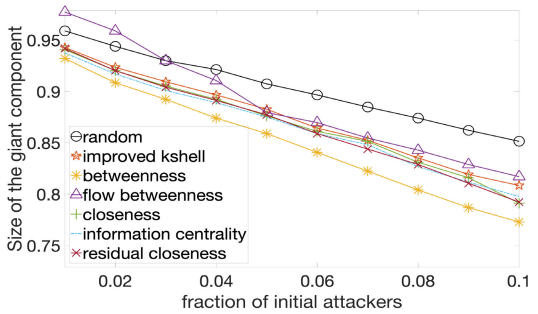
(d) *Iterative point centrality*: Infectious attacks with kshell, mixed degree decomposition, neighborhood coreness, eigenvector, katz, and pagerank in an undirected network.



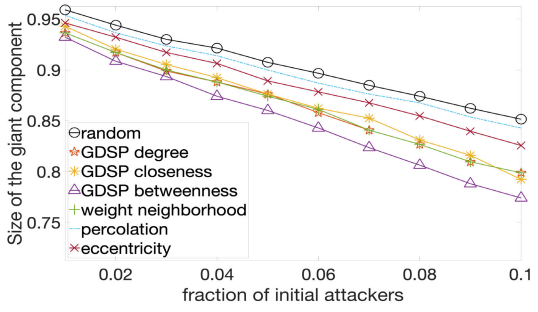
(e) *Iterative point centrality*: Infectious attacks with contribution diffusion centrality subgraph, dynamic influence, and cumulative nomination in an undirected network.



(f) *Iterative point centrality*: Infectious attacks with hubs, authorities, leaderrank, SALSA authorities/hubs in a directed network.



(g) *Global point centrality*: Infectious attacks with improved kshell, flow-betweenness, and closeness, information centrality, residual closeness in an undirected network.



(h) Infectious attacks with improved kshell, flow-betweenness, and closeness, information centrality, residual closeness in an undirected network.

FIGURE 9. The size of the giant component after removing the initial *infectious* attacker nodes based on the point centrality metrics in the undirected EU Email Network and the directed Rocketfuel Social Network where the random node removal is included as a baseline model. (a)-(c) are for *local* point centrality metrics, (d)-(f) are for *iterative* point centrality metrics, and (g)-(h) are for *global* point centrality metrics. All results are based on 100 simulation runs to obtain the mean size of the giant component.

(i.e., the URV Email Network in Fig. 8 and the EU Email Network in Fig. 9) is that the EU Email Network is a denser network than the URV Email Network. This can explain why flow betweenness in (g) can significantly perform better than random in the EU Email Network, compared to its performance in the URV Email Network. That is, since a higher network density (with more edges) can increase the impact of infectious attacks, the flow betweenness-based attacks can take an advantage of the network density to increase its effect in compromising other nodes in the network. In addition, higher network density (with more nodes) can also make the performance of targeted attacks less distinctive because the opportunities for infection are more relevant than the marginal benefits of optimizing the selection of initial attackers.

Effect of Removing the Single Top-Ranked Node: Fig. 10 shows the effect of point centrality-based targeted attacks in the undirected networks (i.e., EU Email Network and URV Email Network) and directed networks (i.e., UCI Social Network and Rocketfuel Network) in terms of the size of the giant component as an indicator of the network resilience when the single top-ranked node based on a given metric is selected as an infectious attacker. The trends are very similar to Fig. 10 in terms of the performance under different networks. Repeating the trends observed in Fig. 10, the effect of targeted attacks based on point centrality metrics is greater (i.e., smaller size of the giant component) in the sparse URV Email Network than in the dense EU Email Network. It is not surprising that the dense network can absorb the impact of removing nodes and better maintain a connected network. However, interestingly, in directed networks, the sparsity of the directed Rocketfuel Network can mitigate the infection process, leading to a larger size of the giant component while the higher density of the UCI Social Network allows attacks to more easily spread.

Effect of Varying the Fraction of Initial, Infectious Attackers on the Mean Fraction of Nodes Infected by a Single, Initial Attacker: Figs. 11 and 12 show the mean fraction of nodes infected by a single, initial attacker when the fraction of initial attackers vary from 0.001 to 0.01 with an increment of 0.01 using 38 point centrality metrics to determine the initial selection under the undirected EU Email Network and the directed Rocketfuel Network. Note that 'the mean fraction of infected nodes' is to measure the attack impact introduced by an initial attacker. This will allow us to investigate how many other nodes a single initial attacker has compromised on average.

Most metrics evaluated in this work showed higher rates of infection spread per initial attacker. However, some metrics, such as flow betweenness, clustering coefficient, diffusion centrality, mixed degree decomposition, and SALSA authorities, showed lower rates per initial attacker. Note that an attack resulting in a smaller size of the giant component does not necessarily mean there are more infected nodes because there may exist many uninfected nodes in smaller components. Conversely, lower infection rates due to a given centrality-based selection does not imply that the network is resilient to that particular attack.

TABLE 9. Relative Graph Centrality (RGC) values of 10 GC metrics under infectious attacks in the undirected network datasets (i.e., EU email network and URV email network).

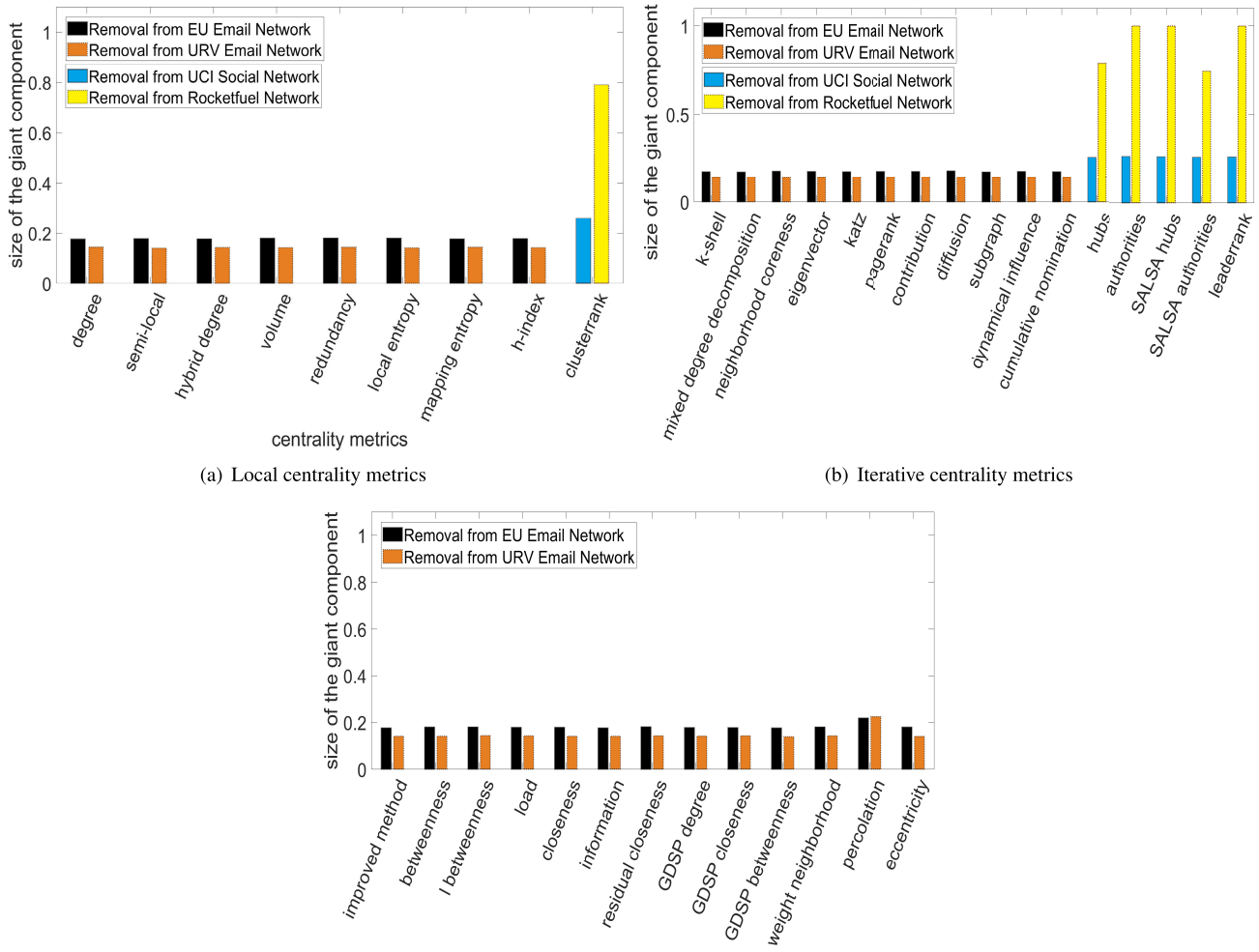
Dataset	EU Email Network		URV Email Network	
% of node removal	30%	70%	30%	70%
Size of the giant component (S_g)	∼0.32	∼0.12	∼0.29	∼0.07
distance-based	0.8833	0.9835	0.8902	0.9929
degree-based	0.8165	0.9433	0.7096	0.8761
k-component	0.302	0.5652	0.3525	0.6977
local assortativity	0.0806	0.2425	0.2206	0.7063
graph curvature	0.1808	0.2345	0.2670	0.3438
average clustering	0.2271	0.4278	0.3780	0.6668
betweenness-based	-0.0049	-0.2095	-1.3418	-1.3036
flow betweenness	-0.1582	-0.3101	-0.5204	-1.0713
closeness-based	0.0194	0.0426	-0.1471	0.1523
degree assortativity	0.0775	-0.1900	0.1608	0.3438

2) UNDER GRAPH CENTRALITY-BASED TARGETED, INFECTIOUS ATTACKS

Table 9 shows the RGC values of the graph centrality (GC) metrics when random infectious attacks are performed. Again, the network is seeded with 30% or 70% of infected nodes and the results are for the two undirected networks, including the EU Email Network and URV Email Network. Due to the infectious nature of this attack, the size of the giant component is observed to be smaller compared to that under non-infectious attacks. But similar to what we observed in Table 8, some GC metrics (e.g., the top 6 GC metrics in Table 8) show a similar tendency with decreasing GC under a graph with a smaller size of the giant component. However, other GC metrics (e.g., the bottom 4 GC metrics in Table 8) do not show a consistent trend. For example, for degree assortativity, the size of GC decreases in the dense EU Email Network while it increases in the sparse URV Email Network. In addition, GC does not always keep increasing or decreasing depending on the size of the giant component even for the same network, as observed in the closeness-based metric. Therefore, the scale of some GC metrics can be used to predict the size of the giant component.

3) UNDER GROUP SELECTION CENTRALITY-BASED TARGETED, INFECTIOUS ATTACKS

Fig. 13 shows the sizes of the giant component in both undirected networks (EU Email Network and URV Email Network) when a set of groups (i.e., a set of 10 nodes) selected based on a given group selection metric are removed as targeted, infectious attacks. Compared to the results under non-infectious attacks in Fig. 7, under infectious attacks, the results are more interesting. For a less dense network like the URV network, the effect of the four metrics on the size of the giant component is similar although



(a) Local centrality metrics

(b) Iterative centrality metrics

(c) Global centrality metrics

FIGURE 10. **The size of the giant component after removing a single top-ranked node as an initial infectious attacker based on the given centrality metric in both undirected networks (i.e., EU Email Network and URV Email Network) and directed networks (i.e., UCI Social Network and Rocketfuel Network).**

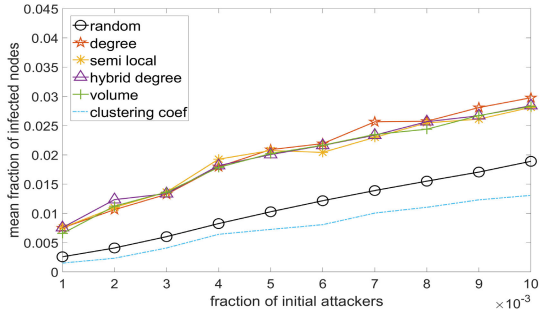
the degree discount seems to be the best selection strategy. However, under the denser network like the EU Email Network, the degree punishment strategy outperforms the others because high network density mitigates the effect of the penalty. From this observation, we found that under infectious attacks, higher network density can significantly mitigate the effect of the targeted attacks. If a network is not sufficiently dense, regardless of what metric is used to select targets to attack, the network can more easily collapse. Thus, it is more important to select the right group selection metric for developing more powerful attacks under dense networks than under sparse networks.

D. RUNNING TIME ANALYSIS

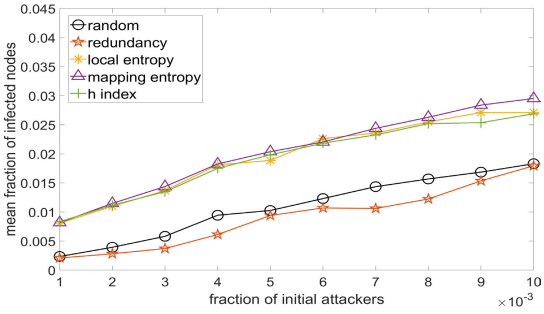
1) RUNNING TIME OF POINT CENTRALITY METRICS

Fig. 14 shows the running time in $\log_{10}$ sec. to show the efficiency of the point centrality metrics surveyed in this work using the undirected URV Email Network and the UCI Social Network and undirected EU Email Network and the Rocketfuel Network, respectively. From Fig. 14, we made the following observations:

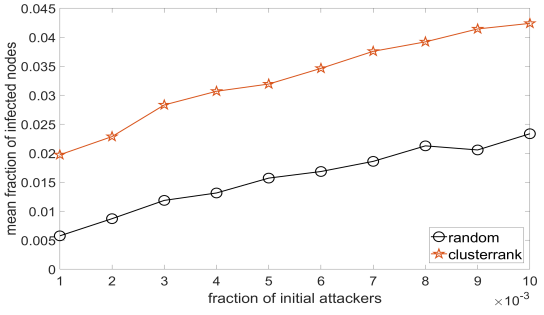
- Degree, pagerank, and GDSP degree exhibit the best efficiency in local, iterative, and global centrality metrics, respectively. This explains well why simple degree-based or similar centrality metrics have been dominantly used in practice based on their efficiency in calculation.
- We observed relatively slow running times in clusterrank, hybrid degree, neighborhood coreness, SALSA hubs, SASA authorities, and percolation centrality metrics. Although these metrics offer certain useful features in capturing insightful centrality concepts in terms of power or influence, their slow running time may not be attractive particularly in sizable or resource-constrained, distributed environments.
- When we compare the performance of each centrality metric under those two sets of network topologies (i.e., URV/UCI and EU/Rocketfuel networks), we found



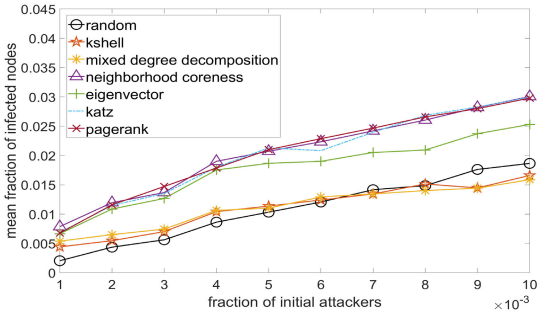
(a) *Local point centrality*: Infectious attacks with degree, semi-local, hybrid degree, volume, and clustering coefficient in an undirected network.



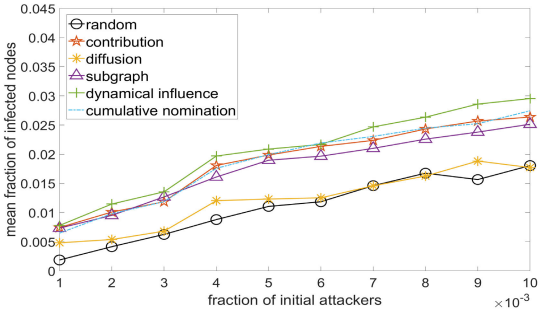
(b) *Local point centrality*: Infectious attacks with redundancy, local entropy, mapping entropy, and h-index in an undirected network.



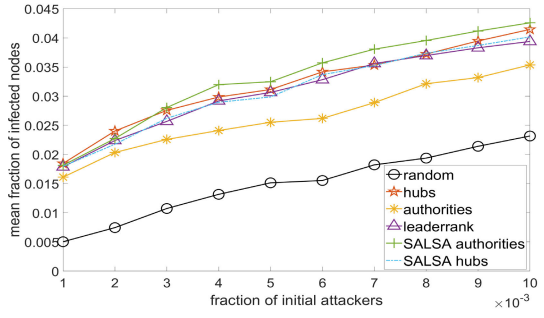
(c) *Local point centrality*: Noninfectious attacks with clusterrank in a directed network.



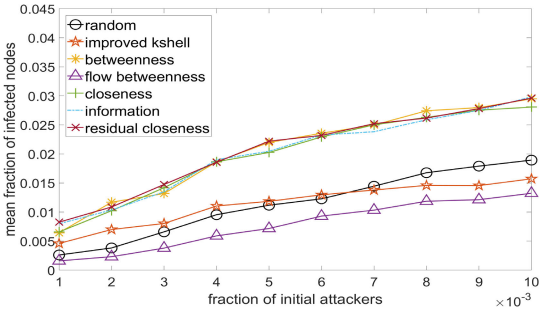
(d) *Iterative point centrality*: Infectious attacks with kshell, mixed degree decomposition, neighborhood coreness, eigenvector, katz, and pagerank in an undirected network.



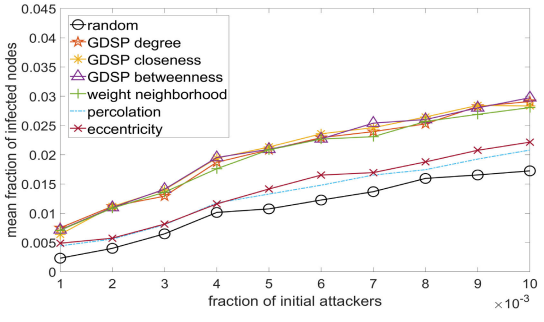
(e) *Iterative point centrality*: Infectious attacks with contribution diffusion centrality subgraph, dynamic influence, and cumulative nomination in an undirected network.



(f) *Iterative point centrality*: Infectious attacks with hubs, authorities, leaderrank, SALSA authorities/hubs in a directed network.

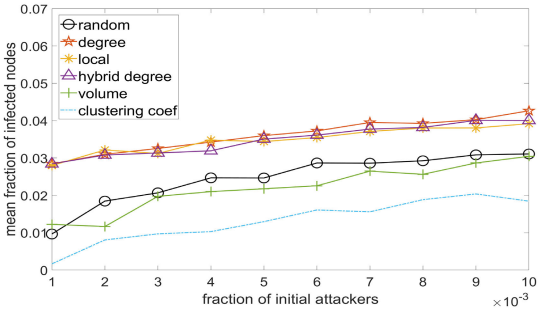


(g) *Global point centrality*: Infectious attacks with improved kshell, betweenness, and closeness, information centrality, residual closeness in an undirected network.

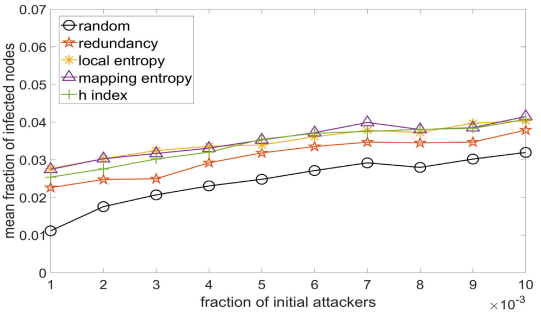


(h) *Global point centrality*: Infectious attacks with GDSP-degree, GDSP-closeness, GDSP-betweenness, weight neighborhood, percolation, and eccentricity in an undirected network.

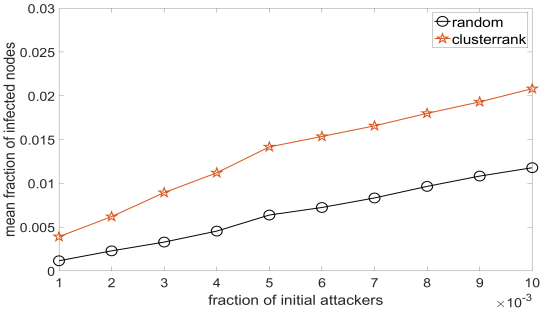
FIGURE 11. Mean fraction of infected nodes after infectious, initial targeted attackers are selected from 0.001 to 0.01 with the increment of 0.01 based on the centrality metrics in the undirected URV Email Network and the directed UCI Social Network. All results are shown based on 100 simulation runs to obtain the mean size of the giant component.



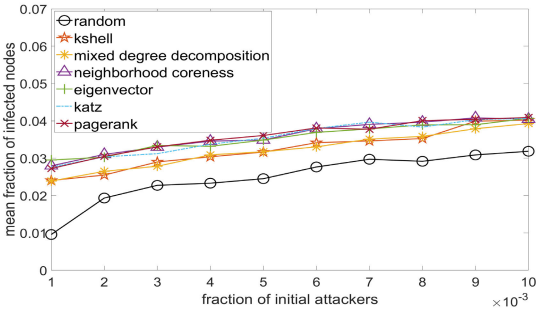
(a) *Local point centrality*: Infectious attacks with degree, semi-local, hybrid degree, volume, and clustering coefficient in an undirected network.



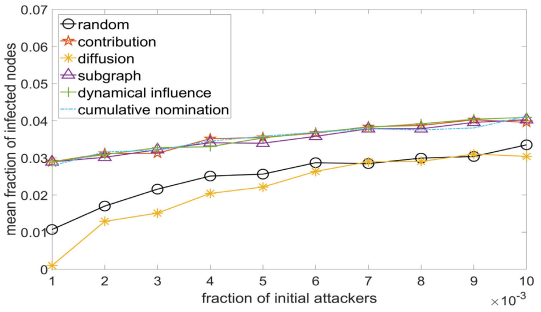
(b) *Local point centrality*: Infectious attacks with redundancy, local entropy, mapping entropy, and h-index in an undirected network.



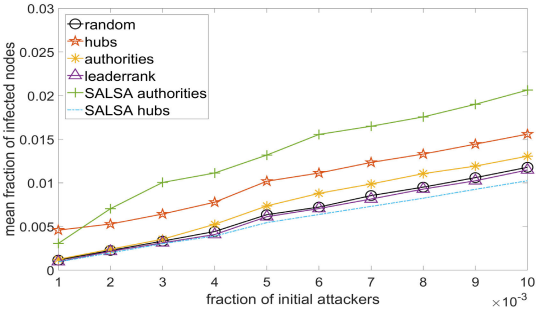
(c) *Local point centrality*: Noninfectious attacks with clusterrank in a directed network.



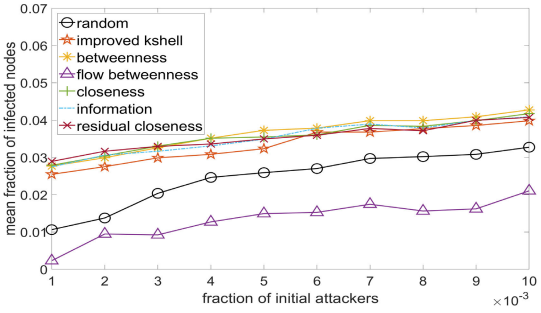
(d) *Iterative point centrality*: Infectious attacks with kshell, mixed degree decomposition, neighborhood coreness, eigenvector, katz, and pagerank in an undirected network.



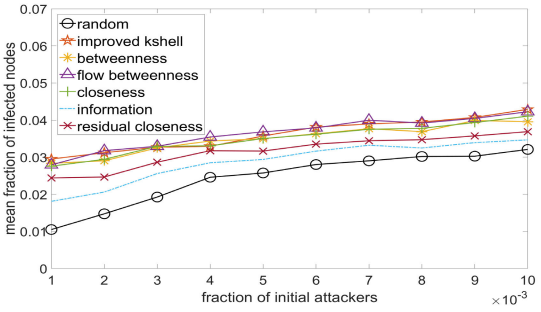
(e) *Iterative point centrality*: Infectious attacks with contribution diffusion centrality subgraph, dynamic influence, and cumulative nomination in an undirected network.



(f) *Iterative point centrality*: Infectious attacks with hubs, authorities, leaderrank, SALSA authorities/hubs in a directed network.

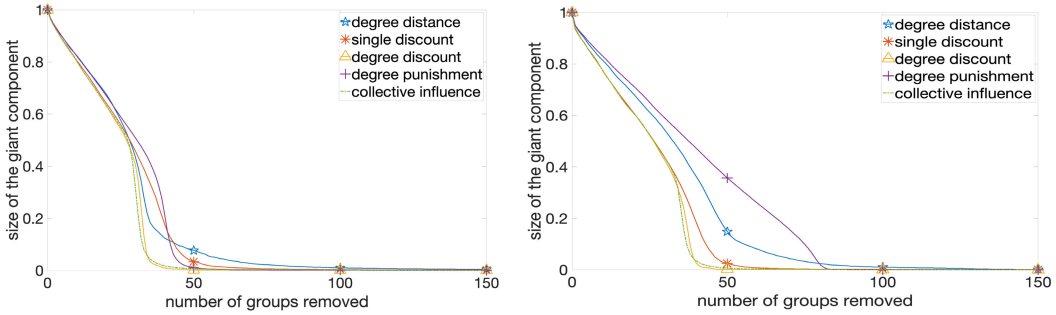


(g) *Global point centrality*: Infectious attacks with improved kshell, betweenness, and closeness, information centrality, residual closeness in an undirected network.



(h) *Global point centrality*: Infectious attacks with GDSP-degree, GDSP-closeness, GDSP-betweenness, weight neighborhood, percolation, and eccentricity in an undirected network.

FIGURE 12. Mean fraction of infected nodes after infectious, initial targeted attackers are selected from 0.001 to 0.01 with the increment of 0.01 based on the centrality metrics in the undirected EU Email Network and the directed Rocketfuel Social Network. All results are shown based on 100 simulation runs to obtain the mean size of the giant component.



(a) Under infectious attacks in the URV Email Network (b) Under infectious attacks in the EU Email Network

FIGURE 13. The size of the giant component after removing a set of infectious initial attackers based a given group selection metrics in the two undirected network datasets, which are the EU Email Network URV Email Network.

there are only slight differences in the performance order. This is because the characteristics of a network dataset affect each centrality metric's running time. However, the trends are similar since the performance order is still dependent on the inherent complexity of each metric.

2) RUNNING TIME OF GRAPH CENTRALITY METRICS

Fig 15 shows the running time of the graph centrality (GC) metrics per simulation run on the undirected URV and EU Email Networks. We found most k-metrics, except k-core, are fairly slow while common metrics such as degree-based metrics are faster, which could be a reason of its popular utilization in various domain applications. However, there is no clear relationship between algorithmic complexity and the nature of the GC metrics, such as local or global metrics, in the process of their calculation. Compared to the results in the URV Email network (see (a), (c), and (e), the results under undirected EU Email Network (see (b), (d), and (f) show a slightly different performance order. However, the overall trend is similar. We observed that there is no relationship between algorithmic complexity and local or global centrality nature in the GC metrics.

3) RUNNING TIME OF GROUP SELECTION CENTRALITY METRICS

Fig. 16 shows the running time of the group selection metrics per simulation round. We found that the degree distance is more expensive than other metrics that are the enhanced versions to improve the complexity of the degree distance using heuristic methods. We also observed that there is a longer running time for calculating the metrics using the URV Email Network than using the EU Email Network. Even though the URV Email Network has more nodes than the EU Email Network, the EU Email Network has five times higher network density (i.e., more edges) than the URV Email Network. This implies that the complexity of a group selection centrality is more affected by node density rather than network density.

VIII. APPLICATIONS OF CENTRALITY METRICS IN VARIOUS NETWORK TYPES

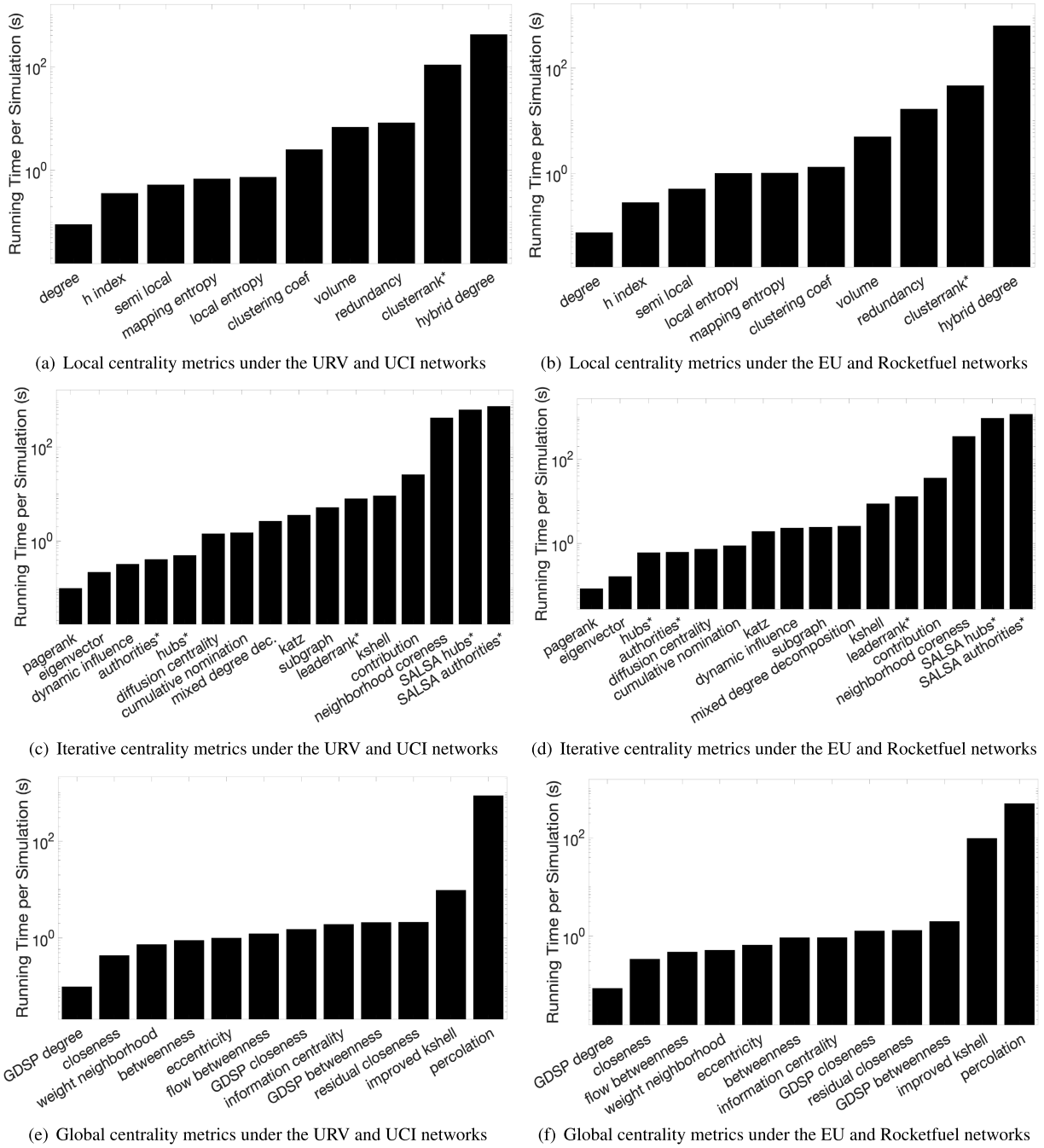
In this section, we give an overview of how centrality metrics have been applied in various types of networks, including social networks, contact networks, communication networks, geographic networks, and biological networks.

A. SOCIAL NETWORKS

1) INFORMATION DIFFUSION

This problem involves determining the initial set of nodes that efficiently propagates information throughout the network. Kim and Yoneki [96] and Kim *et al.* [184] investigated this selection process under different information diffusion strategies. They found that when the initial set of seed propagators are high-degree nodes, then the choice of neighboring nodes to spread the information does not affect the long-term propagation significantly.

Network structure features, such as network topology, node in-degree, out-degree, edge weight, and clustering coefficient, have also been considered in studies of false information propagation [185]–[188]. Cho *et al.* [185] built an uncertainty-based subjective opinion model using a belief model, called Subjective Logic. They developed different types of agents that can propagate false information intentionally (i.e., disinformers) and mistakenly (i.e., misinformers), where true information is also propagated to counter the false information. The authors investigated the effect of different types of centrality metrics used in the selection of sources propagating the false or true information. Kumar *et al.* [186] developed feature sets including network features to identify hoaxes in Wikipedia, including the network centrality measures to represent the relation between the references of the article in the Wikipedia hyperlink network. Ratkiewicz *et al.* [187] built a 'Truthy' system to enable the detection of 'astroturfing' (fake grass root campaigning with hidden sponsors) on Twitter. Wu *et al.* [188] summarized false information spreader detection based on different network topologies. The authors examined how the so-called 'forceful' individuals (not changing their opinions) can affect information diffusion depending on how they are



(a) Local centrality metrics under the URV and UCI networks

(b) Local centrality metrics under the EU and Rocketfuel networks

(c) Iterative centrality metrics under the URV and UCI networks

(d) Iterative centrality metrics under the EU and Rocketfuel networks

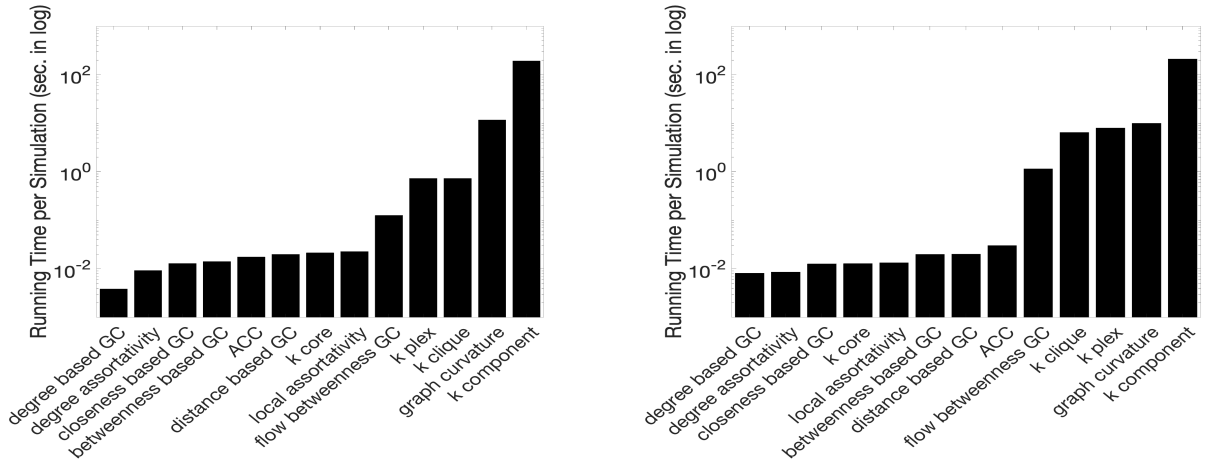
(e) Global centrality metrics under the URV and UCI networks

(f) Global centrality metrics under the EU and Rocketfuel networks

FIGURE 14. **Simulation running time (in $\log_{10}$ sec.) of the point centrality metrics applied to the undirected networks (i.e., URV and EU Email Networks) and the directed networks (i.e., UCI and Rocketfuel Networks). Note that centrality metrics that can be only shown in directed networks are indicated with *.**

connected with other individuals or when they bridge multiple communities.

Kimura *et al.* [189], [190] considered the problem of identifying the most influential nodes in a large-scale social network as a combinatorial optimization problem. The authors proposed an efficient greedy algorithm based on bond percolation and graph theory and demonstrated its superior performance over conventional methods in terms of computational cost. Tang *et al.* [191] investigated an email dataset as a dynamic, social network in order to study dynamic interactions using a proposed 'temporal centrality metric.' Particularly, the authors measured information dissemination using the centrality metric and examined the role of 'information mediators' to better understand the dynamics of the social network and accurately identify central people compared to only using conventional, static centrality metrics. Kandhway and Kuri [192] used an epidemic model to maximize information diffusion for a certain period of



(a) Under the URV email network

(b) Under the EU email network

FIGURE 15. Simulation running time in sec. (in log scale) for the graph centrality metrics applied to the undirected URV and EU Email Networks.

campaign running in a social network. The authors proposed an optimal control framework that can maximize the information diffusion using controls (i.e., advertisement) over the campaign period. They examined the effect of various types of centrality metrics for initial spreaders on information diffusion.

2) INFLUENCE MAXIMIZATION

Bae and Kim [120] focused on classifying the ability of influential nodes in order, avoiding the assignment of multiple nodes to the same order, using neighborhood coreness centrality. This measure is closely related to epidemic models in that higher influence implies a broader scope of epidemic spreading. Bian *et al.* [148] adopted the SI (Susceptible-Infected) model to identify influential nodes spreading disease in complex networks by using the Analytical Hierarchy Process (AHP) decision making strategy that combines different centrality metrics which typically include degree, closeness, and betweenness. Chen *et al.* [94] introduced a semi-local centrality metric and used a modified version of the SIR model to verify its correctness. The difference compared to the original SIR model is that not every neighbor of an infected node will be infected with a particular propagation probability but rather one neighbor is chosen randomly and infected with certainty. Bavelas *et al.* [3] indicated that a centrality position in small groups influences the perceptions of leadership (as well as morale). Newman [138] demonstrated how random-walk betweenness is a better measure than the degree in the Florentine families intermarriage network [193]. Mochalova and Nanopoulos [194] examined the relationships between the influence of key members and the attitude the remaining members have towards information and how the relationship impacts information diffusion and its outcome.

A key goal in marketing or information diffusion research is to identify influentials, a small set of nodes that can significantly affect a large portion of their network. Watts and Dodds [195] questioned this hypothesis and studied if the size of influence cascades is truly caused by the information propagated from the influentials. Saito *et al.* [196] studied the identification of super-mediators, nodes playing a significant role in receiving or passing information between other nodes in social networks. Goya *et al.* [197] studied a fundamental problem in terms of where or how the input parameters to study an influence model in social networks can be obtained.

Identifying influential nodes in complex networks has been substantially studied by improving existing centrality metrics or taking more comprehensive approaches. Liu *et al.* [198] proposed a generalized weighted gravity model, called *generalized mechanics model* (GMM), by considering global information and local information. Wen *et al.* [199] proposed multi-local dimension (MLD) based on the fractal property to identify vital spreaders in complex networks. MLD is considered as a more general method, such as some existing centrality metrics. Unlike classical centrality metrics, lower MLD indicates higher influence. Li *et al.* [200] also proposed a generalized gravity model measuring local information from both the local clustering coefficient and the degree of each node to identify influential nodes in complex networks. In order to improve the existing centrality-based approach to identify vital nodes in complex networks, Zhao *et al.* [201] adopted the Kullback-Leibler divergence to measure the structural similarity of nodes and updated the PageRank of nodes based on similarity.

3) INFLUENCE MINIMIZATION

Kimura *et al.* [202] solved an influence minimization problem by blocking a limited number of links that spread false information or rumors, where betweenness and out-degrees are used to identify links or nodes to remove. This study found that removing high out-degree nodes is not necessarily effective compared to blocking a limited number of links to maximize the containment. Dey and Roy [203] also

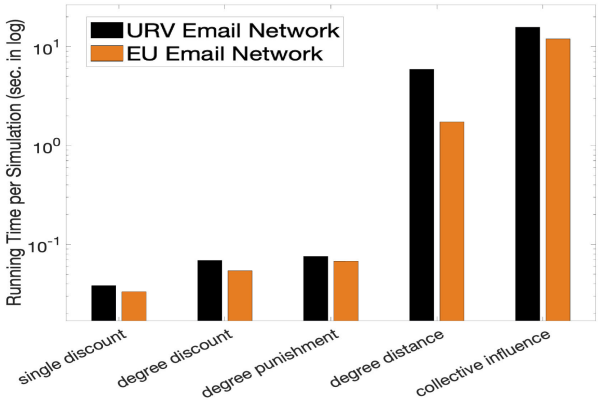


FIGURE 16. Simulation running time in sec. (in log scale) for the group selection metrics applied to the two undirected URV Email Network and EU Email Network. We used $d_{td} = 4$ for the degree distance metric.

studied what nodes to block in order to minimize information propagation. This work used betweenness, edge betweenness, degree, and closeness to block influential nodes. Similarly, Yao *et al.* [204] solved the same problem but by blocking a limited number of nodes where the centrality metrics considered are out-degree and betweenness. Luo *et al.* [205] proposed an algorithm that identifies a set of critical nodes to minimize disinformation in time-varying online social networks. The authors conducted a comparative performance analysis and demonstrated that their proposed algorithm outperforms a centrality-based heuristic counterpart, particularly using degree and closeness.

4) BEHAVIOR ADOPTION FOR MARKETING

Centrality metrics have been also studied as a way to identify initial target populations as a marketing strategy. In adopting technological innovations or purchasing some products, word-of-mouth processes are also modeled using information diffusion models [206]. In particular, as marketing tools, what population to focus on advertising is a major concern, wherein centrality metrics are adopted to identify the target populations [171]. Many marketing applications aimed to leverage social networks or media by targeting populations using simple centrality metrics, such as degrees [207], [208], betweenness [208], [209], or closeness [209], [210].

To study the spreading process of technology adoption, various information maximization algorithms have been proposed and applied to investigate the effect of word of mouth in markets, or game-theoretic strategies [171]. Kempe *et al.* [171] showed that the influence maximization problem is NP-hard and many heuristics or greedy algorithms to solve this problem can provably guarantee a solution to within 63% of the optimal solution, with performance guarantees close to $1 - 1/e$.

5) COMMUNITY DETECTION

Nikolaev *et al.* [211] developed a variant of entropy centrality to understand 'the entropy of flow destination' in networks and showcased how the new entropy centrality is more effective for community detection applications than the original entropy centrality. Jiang *et al.* [212] proposed an efficient centrality measure, called K-rank, designed for selecting the top-K nodes with the highest centrality. The top K nodes are used as the initial seeding nodes and updated based on K-means iterations. The authors applied the K-rank to derive a directed, weighted network for detecting overlapping communities.

B. CONTACT NETWORKS

Christley *et al.* [213] identified the risk of disease infection of nodes using centrality metrics, such as degree, random-walk betweenness, shortest-path betweenness, and farness. Dekker [214] also used six different centrality metrics, including degree, betweenness, two types of closeness, distance-based centrality, and eigenvector centrality in order to identify the super spreaders of infectious diseases. Bell *et al.* [215] investigated the co-relationships between various types of centrality metrics and their variants, such as degree, betweenness, closeness, eigenvector centrality, information centrality, and power prestige. Gomez *et al.* [216] studied high-risk hosts for emerging infectious diseases based on various centrality metrics (e.g., strength, degree, betweenness, closeness, eigenvector centrality) for their control and surveillance.

C. COMMUNICATION NETWORKS

Centrality metrics have also been used to make decisions to solve various problems in communication networks. Centrality metrics have been used to select critical nodes to prevent or mitigate computer virus or malware spreads. Newman *et al.* [14] conducted an empirical study investigating the email network structure to examine what nodes can significantly contribute to spreading computer viruses. Kim [17] measured the risk of websites exposing security vulnerability (e.g., malware, fake infectious sites) based on degree, betweenness, eigenvector, and closeness.

Albert *et al.* [9] showed scale-free networks, following a power-law degree distribution, are highly robust to random attacks while highly vulnerable to targeted attacks on high degree nodes. Holme *et al.* [15] also investigated the network resilience in complex networks when targeted attacks are applied based on degree or betweenness. Yoon *et al.* [16] developed a scalable centrality-based traffic measurement based on software-defined networking functionalities.

D. GEOGRAPHIC NETWORKS

Crucitti *et al.* [147] analyzed spatial networks based on different centrality metrics to characterize the geographic properties of cities as networks. Substituting an undirected graph for urban streets of a city and measuring the different centralities, they presented a spatial distribution of centralities that show the main structures of the city, centric areas, and major routes, depending on the type of measures. Gao *et al.* [217] used betweenness to measure urban traffic flow with GPS-enabled taxi trajectory information in Qingdao, China. This study

TABLE 10. Applications of centrality metrics.

Network Type	Research Problem	Centrality metrics used	Ref. No.
Social Networks	Information diffusion	In-degree; out-degree; clustering-coefficient; temporal centrality; betweenness; closeness; proximity	[185], [96], [184], [189], [192], [190], [186], [187], [188], [191]
	Influence maximization	Coreness; random-walk betweenness; in-degree	[120], [3], [148], [94], [197], [194], [138], [193], [195], [196]
	Influence minimization	Betweenness; out-degree; degree; closeness	[203], [202], [205], [204]
	Behavior adoption for marketing	Degree; betweenness; closeness	[206], [207], [171], [210], [209], [208]
	Community detection	Entropy centrality; K-rank	[212], [211]
Contact Networks	Identification of high-risk hosts or super spreaders	Degree; random-walk betweenness; betweenness; shortest-path betweenness; farness; closeness, distance-based centrality; eigenvector centrality; information centrality; power prestige; strength	[215], [213], [214], [216]
Communication Networks	Selecting critical nodes to prevent or mitigate computer virus or malware spreads; modeling targeted attackers	In-degree; out-degree; degree; betweenness; eigenvector centrality; closeness centrality	[9], [15], [17], [14], [16]
Geographic Networks	Characterizing the geographic properties of cities as networks	Betweenness; closeness; degree; information centrality	[147], [217], [219], [220], [218]
Biological Networks	Removing critical proteins; identifying central nodes such as pathogen-interacting, cancer, aging, HIV-1 or disease related protein	Degree; betweenness; integration; radiality; Katz status index; PageRank; motif-based centralities; weighted sum of loads eigenvector centrality; subgraph centrality; eigenvector centrality	[127], [221], [62], [63]

demonstrated that betweenness is not necessarily a good metric to measure the traffic flow distributions. The authors suggested combining a network structure with other information, such as different patterns of human activities depending on location, power law distance-decay, and human mobility patterns. Porta *et al.* [218] developed a 'Multiple Centrality Assessment (MCA)' framework that uses centrality metrics to understand why the current design features of a city do not attract more people or increase social life. This work used closeness, betweenness, straightness (or degree), and information centrality to understand the current attractiveness of the city. Guimera *et al.* [219] examined the impact of a city's global role based on degree and betweenness. They found that a city's betweenness is more closely related to the city's global role with intercommunity and intracommunity connections than the city's degree centrality. Li *et al.* [220] examined how the centrality of each shipping area, with 25 geographical areas, plays a key role in changing the centrality of the global shipping networks (GSNs) during the years 2011-2012. This study used degree, betweenness, and closeness as centrality metrics to analyze the dynamics of the GSNs.

E. BIOLOGICAL NETWORKS

Estrada *et al.* [127] used centrality to study the removal of proteins from the yeast *S. cereviciae*. The lethality of protein removal has been shown to correlate with the degree of the protein. Jeong *et al.* [221] conducted an experiment of arranging proteins in order of the degree they have and testing the robustness of a network after each protein has been removed. Koschützki1 and Schreiber [62] analyzed the structure of gene regulatory networks based on the ranks of nodes, which are measured by centrality metrics. They used degree, betweenness, integration, radiality, Katz status index, PageRank, and various types of motif-based centralities. Karabekmez and Kirdar [61] proposed a new centrality metric called the *weighted sum of loads eigenvector centrality* (WSL-EC) in order to identify critical nodes in biological networks. The examples are to identify central nodes, such as pathogen-interacting, cancer, aging, HIV-1 or disease-related proteins, proteins involved in immune system processes, and auto-immune diseases in the human interactome. Mistry *et al.* [63] developed a new centrality metric to predict central and critical genes and proteins based on a protein-protein interaction network. The proposed centrality metric considers both the amount of a protein's interaction and the gene coexpression values of genes.

In Table 10, we summarized what centrality metrics have been used in various network types based on our discussions in this work. Although our discussions on the applicability of centrality metrics are limited, this table shows a trend of what centrality metrics have been substantially utilized in contact and biological networks compared to other network domains. Despite a large volume of centrality metrics studied in the literature (see Sections III, IV, and V), we can clearly observe that the uses of centrality metrics have been mostly limited to several common centrality metrics, such as degree (including in/out-degree), betweenness, closeness, and eigenvector centrality.

IX. CONCLUDING REMARKS

In this section, we discuss what we learned from this present study and how to improve the limitations of the existing centrality metrics by suggesting future research directions.

In particular, we surveyed 60 centrality metrics in this work in terms of point, graph, and group selection centrality metrics. We implemented 56 centrality metrics and analyzed

their effect on network resilience based on the size of the giant component when each centrality metric is used to model targeted attacks. We evaluated the performance of each metric under two undirected real network datasets and two directed real network datasets. In this section, we also discuss some insights learned from the findings obtained from the extensive experimental results.

A. LIMITATIONS, INSIGHTS, AND LESSONS LEARNED

We found limitations of the existing centrality metrics surveyed in this work, learned lessons, and obtained the insights from them as follows:

- The meaning of centrality is not only limited to how a node is connected to other nodes, but also implies how actively the node communicates to each other and how it can control or influence other nodes in their centrality or vulnerability. In brief, node centrality determines influence in terms of *connectivity*, *communicability*, and *controllability* in a given network. However, node connectivity is not commonly aligned with the capacity to deal with traffic (e.g., communicability) because nodes with high connectivity are often congested.
- Centrality metrics can be applied in various disciplines with different purposes. In addition, there is a rich volume of centrality metrics available and usable for various design goals. For example, we may want to investigate how to balance traffic loads, how to set edges between nodes to make a network robust against faults or attacks, what types of targeted attacks to develop, how to identify vital nodes based on various criteria, or what is the most least) influential or vulnerable node in a given network.
- We investigated the effect of each centrality metric on network resilience in terms of the size of the giant component. We found that if a centrality metric measures how well a node is connected with its close neighborhood (i.e., locally well connected), its impact upon removing the node with high centrality tends to be limited. For example, removing nodes with high clustering coefficient or volume centrality is not as severe as the random removal of nodes in network resilience (i.e., the size of the giant component). However, if the centrality metric refers to how well the node is globally linked with other nodes which may belong to another cluster of the network (e.g., another community), when the node fails, the network is highly impacted by the node's failure.
- We found that when an attack using a given centrality metric is non-infectious, what metric to choose is highly critical because the effect of a different centrality metric can be vastly different. However, when the attack is infectious, using different centrality metrics does not introduce a significantly different impact on network resilience as the infectious attack itself may be powerful. In addition, we found how a node is connected in a given network (i.e., network topology characteristics such as network density) is the most important factor that can influence the network resilience (i.e., a smaller size of the giant component).
- Although a large volume of centrality metrics has been developed so far, only common centrality metrics have been used, such as degree, betweenness, closeness, clustering coefficient, and pagerank, which has been developed several decades ago. Although the degree is a simple metric, other metrics, such as betweenness or clustering coefficient, require high complexity with slow running time. Even if there have been many centrality metrics developed in the 2010s, not many of them have been used in the existing network applications while the metrics developed from the 1970s to the 1990s have commonly been used in the literature.
- Unlike centrality metrics that are applicable in undirected networks, centrality metrics in directed networks may not be appropriate to study their effect on network resilience. This is because even a node's failure with high centrality (e.g., hub, authority, or leaderrank) in sparse networks may not introduce any significant impact where centrality is mainly measured based on in-degree, not out-degree.
- We used the size of the giant component as an indicator to represent network resilience. The size of the giant component is a conventional network resilience metric in the Network Science domain. However, it does not necessarily indicate how many nodes are compromised as a metric to measure system vulnerability in terms of a cybersecurity perspective. Even if the size of the giant component is small, it does not necessarily imply that the network has more compromised nodes because there could be healthy nodes in smaller components of the network.
- We investigated the running time of all centrality metrics surveyed in this work. The overall trend is that centrality metrics tested under directed networks (e.g., SALSA authorities, SALSA hubs, leaderrank, clusterrank) tend to show higher running time than centrality metrics tested under undirected networks. This may be because undirected networks innately have higher connectivity than directed networks. Recall that many centrality metrics rely on a (shortest) path distance between two nodes as part of the metric calculation.
- The running time of each metric is mainly influenced by network size or network/node density. In addition, in some metrics, we optimized the code to expedite the running time while others may not. Therefore, there may be an inaccuracy introduced in the running times of centrality metrics demonstrated in this work. However, we believe that this imperfect code optimization will not significantly affect the order of running time performance of centrality metrics compared in this work.
- Most point centrality metrics are extensions from the notions of the degree of the node or its neighbors

(e.g. semi-local, k-shell, h-index), connections between the neighbors of the node (e.g., Burt's redundancy, clustering coefficient), pathfinding processes involving the node (e.g., betweenness, closeness), or iterative processes between the node and its neighbors (e.g., eigenvector, pagerank). The extensions attempt to capture something missing or ignored in a fundamental metric (e.g., the degree of the node by itself ignores the degree of its neighbors) whereas semi-local centrality aggregates that information and both k-shell and h-index consider threshold effects on that information. New centrality metrics can be considered by supplementing an existing approach with missing information that may be relevant to the particular problem criteria.

- For an insightful comparison of network resilience under infectious attack using different centrality metrics, the infection rate variability is highly dependent on the characteristic of the network (e.g., network or node density or network topology). Infection is spread more easily in a dense network wherein all the nodes are more easily accessible. On the other hand, a sparse network has structural insulation protecting itself from an infectious attack.

B. FUTURE RESEARCH DIRECTIONS

- *More Efficient Centrality Metrics are Needed:* Since there are many centrality metrics that suffice to meet certain tasks but require less complexity (i.e., low running time), we can leverage these or perhaps modify them to enhance their effectiveness for the task (e.g., increasing the effect of removing a node with high centrality) or efficiency (e.g., running time). Some metrics are representatives of a broader meaning of centrality, such as communicability or controllability (e.g., load centrality in Eq. (30)), in addition to simple connectivity. However, their high complexity hinders applicability in various domains.
- *More Meaningful Metrics are Needed to Measure Network Resilience:* The size of the giant component, as a common metric to measure network resilience, does not reflect a broader concept of network resilience. Network resilience can be defined in terms of how *adaptable* a network is to deal with sudden changes or attacks/failures (i.e., adaptability), how *tolerant* the network is to prevent its failure against attacks or failures (i.e., fault tolerance), and how easily *recoverable* the network is from attacks or failures (i.e., recoverability) [222]. As a future work direction, we need to develop metrics that can measure network resilience embracing adaptability, fault tolerance, and recoverability, or other properties based on system requirements.
- *Graph Centrality Metrics Can be Enhanced as a Novel Measure of Network Resilience:* Graph centrality metrics measure certain characteristics of a given network, such as the distances between nodes, connections between neighbors, or redundant paths between nodes. However, as we observed in Tables 8-9, it is not necessarily correlated to the size of the giant component, which is a conventional metric measuring network resilience in some graph centrality metrics. We can improve the existing graph centrality metrics or invent ones that can be used as indicators related to the key properties of network resilience. For example, when a certain graph centrality value is high, it may indicate the network has the ability to easily recover from attacks or failures.
- *Centrality Metrics Embracing a Broader Concept of Influence Need to be Developed:* Although a rich volume of centrality metrics has been explored in the literature, most of them rely on the concept of centrality based on connectivity. However, in reality, being connected with less critical nodes does not introduce a high impact on network resilience, as long as a small set of critical nodes are still kept safe and operating in a reliable manner. In addition, although controllability is one of the key centrality concepts as discussed in Section II-A, not many centrality metrics are developed without explicitly considering a node's controllability over a given network. There should be more efforts to develop centrality metrics that can fully consider its ability to control the network.
- *Enhancement of the Infection Process for Modeling Infectious Attacks:* In the infection process considered in this work, a node is infected with a given probability. If the node is not infected with the probability, we simply assumed that it is immune to the attack and is not infected again. However, in real-world scenarios, various types of attacks are spread out in a network and there is the possibility that a node can be attacked by multiple or different types of attackers, which allows the same node to be infected multiple times easily. Hence, as a future research direction, a more realistic infection process can be considered where an infected node can recover and be reinfected.
- *In-Depth Analysis of Network Resilience Under Various Network Conditions is Important:* Due to the space constraint, we have not demonstrated more sensitivity analyses to investigate the effect of using a different centrality metric under various network conditions in terms of network density (i.e., the number of edges), node density (i.e., the number of nodes in a given area), or the variance in the number of degrees (e.g., for a scale-free network or a random graph). We can take another in-depth analysis of network resilience by using a different centrality metric in order to identify what metric would be more powerful under what network conditions. In addition, more comprehensive, diverse, larger, and real network topologies can be considered to obtain more meaningful findings to provide generalizable guidelines for selecting useful centrality metrics in a given application.

REFERENCES

[1] L. Euler, "Solutio problematis ad geometriam situs pertinentis," *Commentarii Academiae Scientiarum Petropolitanae*, pp. 128–140, 1741.

[2] A. Bavelas, "A mathematical model for group structures," *Hum. Org.*, vol. 7, no. 3, pp. 16–30, Jul. 1948.

[3] A. Bavelas, "Communication patterns in task-oriented groups," *J. Acoust. Soc. Amer.*, vol. 22, no. 6, pp. 725–730, Jun. 1950.

[4] H. J. Leavitt, "Some effects of communication patterns on group performance," *J. Abnormal Social Psychol.*, vol. 46, pp. 38–50, 1951.

[5] L. Katz, "A new status index derived from sociometric analysis," *Psychometrika*, vol. 18, no. 1, pp. 39–43, Mar. 1953.

[6] B. S. Cohn and M. Marriott, "Networks and centres in the integration of indian civilization," *J. Social Res. I*, pp. 1–9, 1958.

[7] A. Shimbel, "Structural parameters of communication networks," *Bull. Math. Biophys.*, vol. 15, no. 4, pp. 501–507, Dec. 1953.

[8] NRC, *Network Science*. Washington, DC, USA: The National Academies Press, 2005. [Online]. Available: https://www.nap.edu/catalog/11516/network-science

[9] R. Albert, H. Jeong, and A.-L. Barabási, "Error and attack tolerance of complex networks," *Nature*, vol. 406, no. 6794, pp. 378–382, Jul. 2000.

[10] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, "Network robustness and fragility: Percolation on random graphs," *Phys. Rev. Lett.*, vol. 85, no. 25, p. 5468, 2000.

[11] R. Cohen, K. Erez, D. Ben-Avraham, and S. Havlin, "Breakdown of the Internet under intentional attack," *Phys. Rev. Lett.*, vol. 86, no. 16, p. 3682, 2001.

[12] A. L. Barabási and M. Pósfai, *Network Science*. Cambridge, U.K.: Cambridge Univ. Press, 2016.

[13] M. Newman, *Networks: An Introduction*. New York, NY, USA: Oxford Univ. Press, 2010.

[14] M. E. J. Newman, S. Forrest, and J. Balthrop, "Email networks and the spread of computer viruses," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 66, no. 3, Sep. 2002, Art. no. 035101.

[15] P. Holme, B. J. Kim, C. N. Yoon, and S. K. Han, "Attack vulnerability of complex networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 65, no. 5, May 2002, Art. no. 056109.

[16] S. Yoon, T. Ha, S. Kim, and H. Lim, "Scalable traffic sampling using centrality measure on software-defined networks," *IEEE Commun. Mag.*, vol. 55, no. 7, pp. 43–49, Jul. 2017.

[17] D. Kim, "Potential risk analysis method for malware distribution networks," *IEEE Access*, vol. 7, pp. 185157–185167, 2019.

[18] M. Omer, A. Mostashari, and U. Lindemann, "Resilience analysis of soft infrastructure systems," *Procedia Comput. Sci.*, vol. 28, pp. 565–574, Jan. 2014.

[19] D. G. Dessavre, J. E. Ramirez-Marquez, and K. Barker, "Multidimensional approach to complex system resilience analysis," *Rel. Eng. Syst. Saf.*, vol. 149, pp. 34–43, May 2016.

[20] D. Dhyani, W. Ng, and S. Bhowmick, "A survey of Web metrics," *ACM Comput. Surv.*, vol. 34, no. 4, pp. 469–503, Dec. 2002.

[21] 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, Jul. 2013.

[22] 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.

[23] K. Das, S. Samanta, and M. Pal, "Study on centrality measures in social networks: A survey," *Social Netw. Anal. Mining*, vol. 8, no. 1, p. 13, Feb. 2018.

[24] M. Ashtiani, A. Salehzadeh-Yazdi, Z. Razaghi-Moghadam, H. Hennig, O. Wolkenhauer, M. Mirzaie, and M. Jafari, "A systematic survey of centrality measures for protein-protein interaction networks," *BMC Syst. Biol.*, vol. 12, no. 1, Dec. 2018.

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

[26] K. Stephenson and M. Zelen, "Rethinking centrality: Methods and examples," *Social Netw.*, vol. 11, no. 1, pp. 1–37, Mar. 1989.

[27] P. Bonacich, "Power and centrality: A family of measures," *Amer. J. Sociol.*, vol. 92, no. 5, pp. 1170–1182, Mar. 1987.

[28] A. Rusinowska, R. Berghammer, H. De Swart, and M. Grabisch, *Relational and Algebraic Methods in Computer Science*. Berlin, Germany: Springer, 2011, pp. 22–39.

[29] N. E. Friedkin, "Theoretical foundations for centrality measures," *Amer. J. Sociol.*, vol. 96, no. 6, pp. 1478–1504, May 1991.

[30] A. Klein, H. Ahlf, and V. Sharma, "Social activity and structural centrality in online social networks," *Telematics Informat.*, vol. 32, no. 2, pp. 321–332, May 2015.

[31] E. Bakshy, I. Rosenn, C. Marlow, and L. Adamic, "The role of social networks in information diffusion," in *Proc. 21st Int. Conf. World Wide Web (WWW)*, New York, NY, USA: ACM, 2012, pp. 519–528, doi: 10.1145/2187836.2187907.

[32] D. S. Sade, "Sociometrics of macaca mulatta I. Linkages and cliques in grooming matrices," *Folia Primatol.*, vol. 18, nos. 3–4, pp. 196–223, 1972.

[33] T. C. Russo and J. Koesten, "Prestige, centrality, and learning: A social network analysis of an online class," *Commun. Educ.*, vol. 54, no. 3, pp. 254–261, Jul. 2005.

[34] D. Knoke and R. S. Burt, *Applied Network Analysis: A Methodological Introduction*. Beverly Hills, CA, USA: Sage, 1983, pp. 195–222.

[35] K. S. Cook, R. M. Emerson, M. R. Gillmore, and T. Yamagishi, "The distribution of power in exchange networks: Theory and experimental results," *Amer. J. Sociol.*, vol. 89, no. 2, pp. 275–305, Sep. 1983.

[36] R. M. Emerson, "Power-dependence relations," *Amer. Sociol. Rev.*, vol. 27, no. 1, pp. 31–41, 1962.

[37] E. Laumann and F. Pappi, "New directions in the study of community elites," *Amer. Sociol. Rev.*, vol. 38, no. 2, pp. 212–230, 1973.

[38] K. Saito, M. Kimura, K. Ohara, and H. Motoda, "Discovery of super-mediators of information diffusion in social networks," in *Discovery Science*. Berlin, Germany: Springer, 2010, pp. 144–158.

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

[40] K. E. Campbell, P. V. Marsden, and J. S. Hurlbert, "Social resources and socioeconomic status," *Social Netw.*, vol. 8, no. 1, pp. 97–117, Mar. 1986.

[41] J. M. McPherson and L. Smith-Lovin, "Women and weak ties: Differences by sex in the size of voluntary organizations," *Amer. J. Sociol.*, vol. 87, no. 4, pp. 883–904, Jan. 1982.

[42] K. Dënes, *Theory of Finite and Infinite Graphs*. Cambridge, MA, USA: Birkhauser, 1990.

[43] D. J. Klein, "Centrality measure in graphs," *J. Math. Chem.*, vol. 47, no. 4, pp. 1209–1223, May 2010.

[44] J. B. Restrepo, H. A. Forero, and C. A. Cardona, "The analysis of chemical engineering process plants and their models represented by networks," in *Chemical Engineering Transactions*, T. G. Walmsley, P. S. Varbanov, R. Su, and J. J. Klemes, Eds., vol. 70. AIDIC, 2018, pp. 79–84.

[45] I. Gutman, "Chemistry and algebra," *ChemInform*, vol. 36, no. 19, pp. 284–285, May 2005.

[46] D. Allman, "Community centrality and social science research," *Anthropol. Med.*, vol. 22, no. 3, pp. 217–233, Sep. 2015.

[47] S. Collins and M. Durington, *Networked Anthropology: A Primer for Ethnographers*. Routledge, 2014. [Online]. Available: https://books.google.com/books?id=J1iFoAEACAAJ

[48] T. Agryzkov, L. Tortosa, J. F. Vicent, and R. Wilson, "A centrality measure for urban networks based on the eigenvector centrality concept," *Environ. Planning B, Urban Analytics City Sci.*, vol. 46, no. 4, pp. 668–689, May 2019.

[49] F. R. Pitts, "A graph theoretic approach to historical geography," *Prof. Geographer*, vol. 17, no. 5, pp. 15–20, Sep. 1965.

[50] A. Mayer, "Online social networks in economics," *Decis. Support Syst.*, vol. 47, no. 3, pp. 169–184, 2009.

[51] W. Souma, Y. Fujiwara, and H. Aoyama, "Complex networks and economics," *Phys. A, Stat. Mech. Appl.*, vol. 324, no. 1, pp. 396–401, 2003.

[52] S. Epskamp, D. Borsboom, and E. I. Fried, "Estimating psychological networks and their accuracy: A tutorial paper," *Behav. Res. Methods*, vol. 50, no. 1, pp. 195–212, Feb. 2018.

[53] T. Kameda, Y. Ohtsubo, and M. Takezawa, "Centrality in sociocognitive networks and social influence: An illustration in a group decision-making context," *J. Personality Social Psychol.*, vol. 73, no. 2, pp. 296–309, 1997.

[54] S. H. Lee, J. Cotte, and T. J. Noseworthy, "The role of network centrality in the flow of consumer influence," *J. Consum. Psychol.*, vol. 20, no. 1, pp. 66–77, Jan. 2010.

[55] P. Bonacich, "Factoring and weighting approaches to status scores and clique identification," *J. Math. Sociol.*, vol. 2, no. 1, pp. 113–120, Jan. 1972.

[56] S. P. Borgatti and M. G. Everett, "A graph-theoretic perspective on centrality," *Social Netw.*, vol. 28, no. 4, pp. 466–484, Oct. 2006.

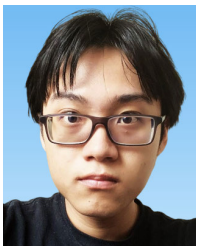
[57] S. Borgatti, "Structural holes: Unpacking Burt's redundancy measures," *Connections*, vol. 20, no. 1, pp. 35–38, 1997.

[58] S. P. Borgatti, "Centrality and AIDS," *Connections*, vol. 18, no. 1, pp. 112–114, 1995.
[59] U. Brandes, "A faster algorithm for betweenness centrality," *J. Math. Sociol.*, vol. 25, no. 2, pp. 163–177, 2001.
[60] L. C. Freeman, S. P. Borgatti, and D. R. White, "Centrality in valued graphs: A measure of betweenness based on network flow," *Social Netw.*, vol. 13, no. 2, pp. 141–154, Jun. 1991.
[61] M. E. Karabekmez and B. Kirdar, "A novel topological centrality measure capturing biologically important proteins," *Mol. BioSyst.*, vol. 12, no. 2, pp. 666–673, 2016.
[62] D. Koschützki and F. Schreiber, "Centrality analysis methods for biological networks and their application to gene regulatory networks," *Gene Regulation Syst. Biol.*, vol. 2, Jan. 2008, Art. no. GRSB.S702.
[63] D. Mistry, M. P. Wise, and J. A. Dickerson, "DiffSLC: A graph centrality method to detect essential proteins of a protein-protein interaction network," *PLoS ONE*, vol. 12, no. 11, pp. 1–25, Nov. 2017.
[64] N. Athanassiou, W. F. Crittenden, L. M. Kelly, and P. Marquez, "Founder centrality effects on the Mexican family firm's top management group: Firm culture, strategic vision and goals, and firm performance," *J. World Bus.*, vol. 37, no. 2, pp. 139–150, Jun. 2002.
[65] L. Kelly, P. M. Lewa, and K. Kamaria, "Founder centrality, management team congruence and performance in family firms: A Kenyan context," *J. Develop. Entrepreneurship*, vol. 13, no. 4, pp. 383–407, Dec. 2008.
[66] V. T. Ho and J. M. Pollack, "Passion Isn't always a good thing: Examining entrepreneurs' network centrality and financial performance with a dualistic model of passion," *J. Manage. Stud.*, vol. 51, no. 3, pp. 433–459, May 2014.
[67] C. Correa, T. Crnovrsanin, and K.-L. Ma, "Visual reasoning about social networks using centrality sensitivity," *IEEE Trans. Vis. Comput. Graphics*, vol. 18, no. 1, pp. 106–120, Jan. 2012.
[68] I. Narayanan, A. Vasan, V. Sarangan, J. Kadengal, and A. Sivasubramaniam, "Little knowledge Isn't always dangerous—Understanding water distribution networks using centrality metrics," *IEEE Trans. Emerg. Topics Comput.*, vol. 2, no. 2, pp. 225–238, Jun. 2014.
[69] Z. Zhang, C. Jiang, S. Guo, Y. Qian, and Y. Ren, "Temporal centrality-balanced traffic management for space satellite networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 4427–4439, May 2018.
[70] H. Zhou, M. Ruan, C. Zhu, V. C. M. Leung, S. Xu, and C.-M. Huang, "A time-ordered aggregation model-based centrality metric for mobile social networks," *IEEE Access*, vol. 6, pp. 25588–25599, 2018.
[71] E. Hafner-Burton and A. Montgomery. (Jan. 2010) *Centrality in Politics: How Networks Confer Power*. [Online]. Available: https://opensiuc.lib.siu.edu/pnconfs_2010/9
[72] D. Lazer, B. Rubineau, C. Chetkovich, N. Katz, and M. Neblo, "The coevolution of networks and political attitudes," *Political Commun.*, vol. 27, no. 3, pp. 248–274, Aug. 2010.
[73] P. R. Miller, P. S. Bobkowski, D. Maliniak, and R. B. Rapoport, "Talking politics on Facebook: Network centrality and political discussion practices in social media," *Political Res. Quart.*, vol. 68, no. 2, pp. 377–391, Jun. 2015.
[74] X.-N. Zuo, R. Ehmke, M. Mennes, D. Imperati, F. X. Castellanos, O. Sporns, and M. P. Milham, "Network centrality in the human functional connectome," *Cerebral Cortex*, vol. 22, no. 8, pp. 1862–1875, Oct. 2011.
[75] G. Saxe, "Network psychiatry: Computational methods to understand the complexity of psychiatric disorders," *J. Amer. Acad. Child Adolescent Psychiatry*, vol. 56, pp. 639–641, Aug. 2017.
[76] Raytheon. (2019). *Network Science Collaborative Technology Alliance (CTA)*. [Online]. Available: http://www.ns-cta.org/ns-cta-blog/
[77] C. J. Colbourn, "Network resilience," *SIAM J. Algebr. Discrete Methods*, vol. 8, no. 3, pp. 404–409, 1987.
[78] W. Najjar and J.-L. Gaudiot, "Network resilience: A measure of network fault tolerance," *IEEE Trans. Comput.*, vol. 39, no. 2, pp. 174–181, Feb. 1990.
[79] T. J. Moore and J.-H. Cho, *Applying Percolation Theory*. Cham, Switzerland: Springer, 2019, pp. 107–133.
[80] D. F. Rueda, E. Calle, and J. L. Marzo, "Robustness comparison of 15 real telecommunication networks: Structural and centrality measurements," *J. Netw. Syst. Manage.*, vol. 25, no. 2, pp. 269–289, Apr. 2017.
[81] D. Santos, A. De Sousa, C. Mas-Machuca, and J. Rak, "Assessment of connectivity-based resilience to attacks against multiple nodes in SDNs," *IEEE Access*, vol. 9, pp. 58266–58286, 2021.
[82] L. C. Freeman, "Centrality in social networks conceptual clarification," *Social Netw.*, vol. 1, no. 3, pp. 215–239, Jan. 1978.
[83] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*, vol. 8. Cambridge, U.K.: Cambridge Univ. Press, 1994.
[84] S. Currarini, M. O. Jackson, and P. Pin, "Identifying the roles of race-based choice and chance in high school friendship network formation," *Proc. Nat. Acad. Sci. USA*, vol. 107, no. 11, pp. 4857–4861, Mar. 2010.
[85] J. Ugander, B. Karrer, L. Backstrom, and C. Marlow, "The anatomy of the Facebook social graph," 2011, *arXiv:1111.4503*. [Online]. Available: http://arxiv.org/abs/1111.4503
[86] C. Wilson, B. Boe, A. Sala, K. P. N. Puttaswamy, and B. Y. Zhao, "User interactions in social networks and their implications," in *Proc. 4th ACM Eur. Conf. Comput. Syst. (EuroSys)*, 2009, pp. 205–218.
[87] R. A. Hanneman and M. Riddle, *Introduction to Social Network Methods*. Riverside, CA, USA: University of California, Riverside, 2005, ch. 10.
[88] U. Brandes, P. Kenis, and D. Wagner, "Communicating centrality in policy network drawings," *IEEE Trans. Vis. Comput. Graphics*, vol. 9, no. 2, pp. 241–253, Apr. 2003.
[89] L. Kleinrock and J. Silvester, "Optimum transmission radii for packet radio networks or why six is a magic number," in *Proc. IEEE Nat. Telecommun. Conf.*, vol. 4, Dec. 1978, pp. 1–4.
[90] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On power-law relationships of the Internet topology," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 29, no. 4, pp. 251–262, Oct. 1999.
[91] R. Albert, I. Albert, and G. L. Nakarado, "Structural vulnerability of the north American power grid," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 69, no. 2, Feb. 2004, Art. no. 025103.
[92] J. E. Hirsch, "An index to quantify an individual's scientific research output," *Proc. Nat. Acad. Sci. USA*, vol. 102, no. 46, pp. 16569–16572, 2005.
[93] E. Garfield, "Citation analysis as a tool in journal evaluation," *Science*, vol. 178, no. 4060, pp. 471–479, 1972. [Online]. Available: https://science.sciencemag.org/content/178/4060/471
[94] 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, pp. 1777–1787, Feb. 2012.
[95] Q. Ma and J. Ma, "Identifying and ranking influential spreaders in complex networks with consideration of spreading probability," *Phys. A, Stat. Mech. Appl.*, vol. 465, pp. 312–330, Jan. 2017.
[96] H. Kim and E. Yoneki, "Influential neighbours selection for information diffusion in online social networks," in *Proc. 21st Int. Conf. Comput. Commun. Netw. (ICCCN)*, Jul. 2012, pp. 1–7.
[97] K. Wehmuth and A. Ziviani, "Distributed assessment of the closeness centrality ranking in complex networks," in *Proc. 4th Annu. Workshop Simplifying Complex Netw. Practitioners (SIMPLEX)*, 2012, pp. 43–48.
[98] R. S. Burt, *Structural Holes: The Social Structure of Competition*. Cambridge, MA, USA: Harvard Univ. Press, Aug. 1995.
[99] D. J. Watts and S. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, Jun. 1998.
[100] K. Anand and G. Bianconi, "Entropy measures for networks: Toward an information theory of complex topologies," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 80, no. 4, Oct. 2009, Art. no. 045102.
[101] T. Nie, Z. Guo, K. Zhao, and Z.-M. Lu, "Using mapping entropy to identify node centrality in complex networks," *Phys. A, Stat. Mech. Appl.*, vol. 453, pp. 290–297, Jul. 2016.
[102] D.-B. Chen, H. Gao, L. Lü, and T. Zhou, "Identifying influential nodes in large-scale directed networks: The role of clustering," *PLoS ONE*, vol. 8, no. 10, pp. 1–10, Oct. 2013.
[103] A. Korn, A. Schubert, and A. Telcs, "Lobby index in networks," *Phys. A, Stat. Mech. Appl.*, vol. 388, no. 11, pp. 2221–2226, Jun. 2009.
[104] L. Lü, T. Zhou, Q.-M. Zhang, and H. E. Stanley, "The H-index of a network node and its relation to degree and coreness," *Nature Commun.*, vol. 7, no. 1, p. 10168, Apr. 2016.
[105] D. Krioukov, F. Papadopoulos, M. Kitsak, A. Vahdat, and M. Boguñá, "Hyperbolic geometry of complex networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 82, no. 3, Sep. 2010, Art. no. 036106.
[106] M. Boguñá, F. Papadopoulos, and D. Krioukov, "Sustaining the Internet with hyperbolic mapping," *Nature Commun.*, vol. 1, no. 1, pp. 1–8, Dec. 2010.
[107] J.-P. Eckmann and E. Moses, "Curvature of co-links uncovers hidden thematic layers in the World Wide Web," in *Proc. Nat. Acad. Sci. USA*, vol. 99, no. 9, pp. 5825–5829, 2002.

[108] M. Keller, "Curvature, geometry and spectral properties of planar graphs," *Discrete Comput. Geometry*, vol. 46, no. 3, pp. 500–525, Oct. 2011.
[109] Y. Higuchi, "Combinatorial curvature for planar graphs," *J. Graph Theory*, vol. 38, no. 4, pp. 220–229, 2001.
[110] O. Knill, "On index expectation and curvature for networks," 2012, *arXiv:1202.4514*. [Online]. Available: http://arxiv.org/abs/1202.4514
[111] Z. Wu, G. Menichetti, C. Rahmede, and G. Bianconi, "Emergent complex network geometry," *Sci. Rep.*, vol. 5, no. 1, pp. 1–12, Sep. 2015.
[112] Y. Ollivier, "Ricci curvature of Markov chains on metric spaces," 2007, *arXiv:math/0701886*. https://arxiv.org/abs/math/0701886
[113] J. Jost and S. Liu, "Ollivier's Ricci curvature, local clustering and curvature-dimension inequalities on graphs," *Discrete Comput. Geometry*, vol. 51, no. 2, pp. 300–322, Mar. 2014.
[114] R. S. Sandhu, T. T. Georgiou, and A. R. Tannenbaum, "Ricci curvature: An economic indicator for market fragility and systemic risk," *Sci. Adv.*, vol. 2, no. 5, May 2016, Art. no. e1501495.
[115] C. Wang, E. Jonckheere, and R. Banirazi, "Interference constrained network control based on curvature," in *Proc. Amer. Control Conf. (ACC)*, Jul. 2016, pp. 6036–6041.
[116] R. Forman, "Bochner's method for cell complexes and combinatorial Ricci curvature," *Discrete Comput. Geometry*, vol. 29, no. 3, pp. 323–374, Feb. 2003.
[117] R. P. Sreejith, K. Mohanraj, J. Jost, E. Saucan, and A. Samal, "Forman curvature for complex networks," *J. Stat. Mech., Theory Exp.*, vol. 2016, no. 6, Jun. 2016, Art. no. 063206.
[118] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse, "Identification of influential spreaders in complex networks," *Nature Phys.*, vol. 6, no. 11, pp. 888–893, Nov. 2010.
[119] A. Zeng and C.-J. Zhang, "Ranking spreaders by decomposing complex networks," *Phys. Lett. A*, vol. 377, no. 14, pp. 1031–1035, Jun. 2013.
[120] J. Bae and S. Kim, "Identifying and ranking influential spreaders in complex networks by neighborhood coreness," *Phys. A, Stat. Mech. Appl.*, vol. 395, pp. 549–559, Feb. 2014.
[121] J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," *J. ACM*, vol. 46, no. 5, pp. 604–632, Sep. 1999.
[122] D. Gibson, J. Kleinberg, and P. Raghavan, "Inferring Web communities from link topology," in *Proc. 9th ACM Conf. Hypertext Hypermedia, Links, Objects, Time Space Structure Hypermedia Syst., Links, Objects, Time Space Structure Hypermedia Syst.*, 1998, pp. 225–234.
[123] S. Brin and L. Page, "The anatomy of a large-scale hypertextual Web search engine," in *Computer Networks and ISDN Systems*. Amsterdam, The Netherlands: Elsevier, 1998, pp. 107–117.
[124] A. J. Alvarez-Socorro, G. C. Herrera-Almarza, and L. A. González-Díaz, "Eigencentrality based on dissimilarity measures reveals central nodes in complex networks," *Sci. Rep.*, vol. 5, no. 1, Dec. 2015, 17095.
[125] J. Bank and B. Cole, "Calculating the Jaccard similarity coefficient with map reduce for entity pairs in Wikipedia," Wikipedia Similarity Team, Tech. Rep., Dec. 2018.
[126] A. Banerjee, A. G. Chandrasekhar, E. Duflo, and M. O. Jackson, "The diffusion of microfinance," *Science*, vol. 341, no. 6144, 2013, Art. no. 1236498. [Online]. Available: http://science.sciencemag.org/content/341/6144/1236498
[127] E. Estrada and J. A. Rodríguez-Velázquez, "Subgraph centrality in complex networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 71, no. 5, May 2005, Art. no. 056103.
[128] L. Lü, Y. Zhang, C. Yeung, and T. Zhou, "Leaders in social networks, the *delicious* case," *PLoS ONE*, vol. 6, no. 6, pp. 1–9, Jun. 2011.
[129] K. Klemm, M. Á. Serrano, V. M. Eguíluz, and M. S. Miguel, "A measure of individual role in collective dynamics," *Sci. Rep.*, vol. 2, no. 1, p. 292, Dec. 2012.
[130] R. Poulin, M.-C. Boily, and B. R. Mâsse, "Dynamical systems to define centrality in social networks," *Social Netw.*, vol. 22, no. 3, pp. 187–220, Jul. 2000.
[131] R. Lempel and S. Moran, "The stochastic approach for link-structure analysis (SALSA) and the TKC effect," *Comput. Netw.*, vol. 33, nos. 1–6, pp. 387–401, Jun. 2000.
[132] J.-G. Liu, Z.-M. Ren, and Q. Guo, "Ranking the spreading influence in complex networks," *Phys. A, Stat. Mech. Appl.*, vol. 392, no. 18, pp. 4154–4159, Sep. 2013.
[133] G. Yan, T. Zhou, B. Hu, Z.-Q. Fu, and B.-H. Wang, "Efficient routing on complex networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 73, no. 4, Apr. 2006, Art. no. 046108.
[134] K.-I. Goh, E. Oh, H. Jeong, B. Kahng, and D. Kim, "Classification of scale-free networks," *Proc. Nat. Acad. Sci. USA*, vol. 99, no. 20, pp. 12583–12588, 2002.
[135] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *Proc. Nat. Acad. Sci. USA*, vol. 99, no. 12, pp. 7821–7826, Apr. 2002.
[136] M. Ercsey-Ravasz and Z. Toroczkai, "Centrality scaling in large networks," *Phys. Rev. Lett.*, vol. 105, no. 3, Jul. 2010, Art. no. 038701.
[137] L. R. Ford and D. R. Fulkerson, *Maximal Flow Through a Network*. Boston, MA, USA: Birkhäuser, 1987, pp. 243–248.
[138] M. E. J. Newman, "A measure of betweenness centrality based on random walks," *Social Netw.*, vol. 27, no. 1, pp. 39–54, Jan. 2005.
[139] K.-I. Goh, B. Kahng, and D. Kim, "Universal behavior of load distribution in scale-free networks," *Phys. Rev. Lett.*, vol. 87, no. 27, Dec. 2001, Art. no. 278701.
[140] U. Brandes, "On variants of shortest-path betweenness centrality and their generic computation," *Social Netw.*, vol. 30, no. 2, pp. 136–145, May 2008.
[141] S. Dolev, Y. Elovici, and R. Puzis, "Routing betweenness bentrality," *J. ACM*, vol. 57, no. 4, pp. 25:1–25:27, Apr. 2010.
[142] U. Brandes and D. Fleischer, "Centrality measures based on current flow," in *STACS*, vol. 3404. Springer, 2005, pp. 533–544.
[143] C. Dangalchev, "Residual closeness in networks," *Phys. A, Stat. Mech. Appl.*, vol. 365, no. 2, pp. 556–564, Jun. 2006.
[144] M. O. Jackson and A. Wolinsky, "A strategic model of social and economic networks," *J. Econ. Theory*, vol. 71, no. 1, pp. 44–74, Oct. 1996.
[145] M. O. Jackson, *Social and Economic Networks*. Princeton, NJ, USA: Princeton Univ. Press, 2010.
[146] N. Tsakas, "On decay centrality," Tech. Rep., 2016.
[147] P. Crucitti, V. Latora, and S. Porta, "Centrality measures in spatial networks of urban streets," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 73, no. 3, Mar. 2006, Art. no. 036125.
[148] T. Bian, J. Hu, and Y. Deng, "Identifying influential nodes in complex networks based on AHP," *Phys. A, Stat. Mech. Appl.*, vol. 479, pp. 422–436, Aug. 2017.
[149] J. Hu, Y. Du, H. Mo, D. Wei, and Y. Deng, "A modified weighted TOPSIS to identify influential nodes in complex networks," *Phys. A, Stat. Mech. Appl.*, vol. 444, pp. 73–85, Feb. 2016.
[150] A. Barrat, M. Barthelemy, R. Pastor-Satorras, and A. Vespignani, "The architecture of complex weighted networks," *Proc. Nat. Acad. Sci. USA*, vol. 101, no. 11, pp. 3747–3752, Mar. 2004.
[151] M. E. J. Newman, "Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 64, no. 1, Jun. 2001, Art. no. 016132.
[152] T. Opsahl, F. Agneessens, and J. Skvoretz, "Node centrality in weighted networks: Generalizing degree and shortest paths," *Social Netw.*, vol. 32, no. 3, pp. 245–251, Jul. 2010.
[153] J. Wang, X. Hou, K. Li, and Y. Ding, "A novel weight neighborhood centrality algorithm for identifying influential spreaders in complex networks," *Phys. A, Stat. Mech. Appl.*, vol. 475, pp. 88–105, Jun. 2017.
[154] M. Piraveenan, M. Prokopenko, and L. Hossain, "Percolation centrality: Quantifying graph-theoretic impact of nodes during percolation in networks," *PLoS ONE*, vol. 8, no. 1, Jan. 2013, Art. no. e53095.
[155] P. Hage and F. Harary, "Eccentricity and centrality in networks," *Social Netw.*, vol. 17, no. 1, pp. 57–63, Jan. 1995.
[156] G. Sabidussi, "The centrality index of a graph," *Psychometrika*, vol. 31, no. 4, pp. 581–603, Dec. 1966. [Online]. Available: https://EconPapers.repec.org/RePEc:spr:psycho:v:31:y:1966:i:4:p:581-603
[157] A. Saxena, R. Gera, and S. R. S. Iyengar, "A faster method to estimate closeness centrality ranking," 2017, *arXiv:1706.02083*. [Online]. Available: http://arxiv.org/abs/1706.02083
[158] J. Nieminen, "On the centrality in a graph," *Scandin. J. Psychol.*, vol. 15, no. 1, pp. 332–336, 1974.
[159] S. B. Seidman and B. L. Foster, "A graph-theoretic generalization of the clique concept," *J. Math. Sociol.*, vol. 6, no. 1, pp. 139–154, Jan. 1978.
[160] N. Tichy, "An analysis of clique formation and structure in organizations," *Administ. Sci. Quart.*, vol. 18, no. 2, pp. 194–208, 1973.
[161] P. W. Holland and S. Leinhardt, "Transitivity in structural models of small groups," *Comparative Group Stud.*, vol. 2, no. 2, pp. 107–124, May 1971.
[162] M. E. J. Newman, "Assortative mixing in networks," *Phys. Rev. Lett.*, vol. 89, no. 20, Oct. 2002, 208701.
[163] R. Noldus and P. Van Mieghem, "Assortativity in complex networks," *J. Complex Netw.*, vol. 3, no. 4, pp. 507–542, Dec. 2015.

[164] M. E. J. Newman, "Mixing patterns in networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 67, no. 2, Feb. 2003.
[165] M. Piraveenan, M. Prokopenko, and A. Y. Zomaya, "Local assortativeness in scale-free networks," *EPL*, vol. 89, no. 4, p. 49901, Feb. 2010.
[166] O. Narayan and I. Saniee, "Large-scale curvature of networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 84, no. 6, Dec. 2011, Art. no. 066108.
[167] M. Gromov, "Hyperbolic groups," in *Essays in Group Theory*. New York, NY, USA: Springer, 1987, pp. 75–263.
[168] E. A. Jonckheere, P. Lohsoonthorn, and F. Ariaei, "Upper bound on scaled gromov-hyperbolic δ," *Appl. Math. Comput.*, vol. 192, no. 1, pp. 191–204, Sep. 2007.
[169] A. Sheikhahmadi, M. A. Nematbakhsh, and A. Shokrollahi, "Improving detection of influential nodes in complex networks," *Phys. A, Stat. Mech. Appl.*, vol. 436, pp. 833–845, Oct. 2015.
[170] W. Chen, Y. Wang, and S. Yang, "Efficient influence maximization in social networks," in *Proc. 15th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2009, pp. 199–208.
[171] D. Kempe, J. Kleinberg, and É. Tardos, "Maximizing the spread of influence through a social network," in *Proc. 9th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, New York, NY, USA, 2003, pp. 137–146.
[172] X. Wang, Y. Su, C. Zhao, and D. Yi, "Effective identification of multiple influential spreaders by DegreePunishment," *Phys. A, Stat. Mech. Appl.*, vol. 461, pp. 238–247, Nov. 2016.
[173] F. Morone and H. A. Makse, "Influence maximization in complex networks through optimal percolation," *Nature*, vol. 524, no. 7563, pp. 65–68, Aug. 2015.
[174] F. Morone, B. Min, L. Bo, R. Mari, and H. A. Makse, "Collective influence algorithm to find influencers via optimal percolation in massively large social media," *Sci. Rep.*, vol. 6, no. 1, pp. 1–11, Jul. 2016.
[175] P. Erdos and A. Rényi, "On the evolution of random graphs," *Pub. Math. Inst. Hungarian Acad. Sci.*, vol. 5, no. 1, pp. 17–60, 1960.
[176] E. Estrada and J. A. Rodríguez-Velázquez, "Subgraph centrality and clustering in complex hyper-networks," *Phys. A, Stat. Mech. Appl.*, vol. 364, pp. 581–594, May 2006. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0378437105012550
[177] E. Estrada and N. Hatano, "Communicability in complex networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 77, no. 3, Mar. 2008, Art. no. 036111.
[178] P. Panzarasa, T. Opsahl, and K. M. Carley, "Patterns and dynamics of users' behavior and interaction: Network analysis of an online community," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 60, no. 5, pp. 911–932, May 2009. [Online]. Available: https://snap.stanford.edu/data/CollegeMsg.html
[179] N. Spring, R. Mahajan, and D. Wetherall, "Measuring ISP topologies with rocketfuel," in *Proc. Conf. Appl., Technol., Archit., Protocols Comput. Commun. (SIGCOMM)*, 2002, pp. 133–145. [Online]. Available: http://networkrepository.com/tech-routers-rf.php
[180] R. A. Rossi and N. K. Ahmed, "The network data repository with interactive graph analytics and visualization," in *Proc. AAAI*, 2015. [Online]. Available: http://networkrepository.com/ia-email-univ.php
[181] A. Paranjape, A. R. Benson, and J. Leskovec, "Motifs in temporal networks," in *Proc. 10th ACM Int. Conf. Web Search Data Mining*, Feb. 2017. [Online]. Available: https://snap.stanford.edu/data/email-Eu-core-temporal.html
[182] M. Alam, D. Yang, J. Rodriguez, and R. Abd-alhameed, "Secure device-to-device communication in LTE—A," *IEEE Commun. Mag.*, vol. 52, no. 4, pp. 66–73, Apr. 2014.
[183] S. Mavoungou, G. Kaddoum, M. Taha, and G. Matar, "Survey on threats and attacks on mobile networks," *IEEE Access*, vol. 4, pp. 4543–4572, 2016.
[184] H. Kim, K. Beznosov, and E. Yoneki, "A study on the influential neighbors to maximize information diffusion in online social networks," *Comput. Social Netw.*, vol. 2, no. 1, Dec. 2015.
[185] J.-H. Cho, S. Rager, J. O'Donovan, S. Adalı, and B. D. Horne, "Uncertainty-based false information propagation in social networks," *ACM Trans. Social Comput.*, vol. 2, no. 2, pp. 1–34, Oct. 2019.
[186] S. Kumar, R. West, and J. Leskovec, "Disinformation on the Web: Impact, characteristics, and detection of Wikipedia hoaxes," in *Proc. 25th Int. Conf. World Wide Web*, Apr. 2016, pp. 591–602.
[187] J. Ratkiewicz, M. Conover, M. Meiss, B. Gonçalves, S. Patil, A. Flammini, and F. Menczer, "Truthy: Mapping the spread of astroturf in microblog streams," in *Proc. 20th Int. Conf. Companion World Wide Web (WWW)*, 2011, pp. 249–252.
[188] L. Wu, F. Morstatter, X. Hu, and H. Liu, *Mining Misinformation in Social Media*. Boca Raton, FL, USA: CRC Press, 2016, pp. 135–162.
[189] M. Kimura, K. Saito, and R. Nakano, "Extracting influential nodes for information diffusion on a social network," in *Proc. 22nd Nat. Conf. Artif. Intell.*, vol. 2. Palo Alto, CA, USA: AAAI Press, 2007, pp. 1371–1376.
[190] M. Kimura, K. Saito, R. Nakano, and H. Motoda, "Extracting influential nodes on a social network for information diffusion," *Data Mining Knowl. Discovery*, vol. 20, no. 1, p. 70, Oct. 2009.
[191] J. Tang, M. Musolesi, C. Mascolo, V. Latora, and V. Nicosia, "Analysing information flows and key mediators through temporal centrality metrics," in *Proc. 3rd Workshop Social Netw. Syst. (SNS)*, 2010, pp. 3:1–3:6.
[192] K. Kandhway and J. Kuri, "Using node centrality and optimal control to maximize information diffusion in social networks," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 47, no. 7, pp. 1099–1110, Jul. 2017.
[193] J. F. Padgett and C. K. Ansell, "Robust action and the rise of the medici," *Amer. J. Sociol.*, vol. 98, no. 6, pp. 1259–1319, 1993.
[194] A. Mochalova and A. Nanopoulos, "On the role of centrality in information diffusion in social networks," in *Proc. ECIS*, 2013.
[195] D. J. Watts and P. S. Dodds, "Influentials, networks, and public opinion formation," *J. Consum. Res.*, vol. 34, no. 4, pp. 441–458, Dec. 2007.
[196] K. Saito, M. Kimura, K. Ohara, and H. Motoda, "Discovery of super-mediators of information diffusion in social networks," in *Discovery Science*. Berlin, Germany: Springer, 2010, pp. 170–184.
[197] A. Goyal, F. Bonchi, and L. V. S. Lakshmanan, "Learning influence probabilities in social networks," in *Proc. 3rd ACM Int. Conf. Web Search Data Mining (WSDM)*. New York, NY, USA: AC, 2010, pp. 241–250.
[198] F. Liu, Z. Wang, and Y. Deng, "GMM: A generalized mechanics model for identifying the importance of nodes in complex networks," *Knowl.-Based Syst.*, vol. 193, Apr. 2020, Art. no. 105464. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0950705119306744
[199] T. Wen, D. Pelusi, and Y. Deng, "Vital spreaders identification in complex networks with multi-local dimension," *Knowl.-Based Syst.*, vol. 195, May 2020, Art. no. 105717. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0950705120301398
[200] H. Li, Q. Shang, and Y. Deng, "A generalized gravity model for influential spreaders identification in complex networks," *Chaos, Solitons Fractals*, vol. 143, Feb. 2021, Art. no. 110456. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0960077920308481
[201] J. Zhao, Y. Song, F. Liu, and Y. Deng, "The identification of influential nodes based on structure similarity," *Connection Sci.*, vol. 33, no. 2, pp. 201–218, Apr. 2021.
[202] M. Kimura, K. Saito, and H. Motoda, "Blocking links to minimize contamination spread in a social network," *ACM Trans. Knowl. Discovery Data*, vol. 3, no. 2, pp. 1–23, Apr. 2009, doi: 10.1145/1514888.1514892.
[203] P. Dey and S. Roy, "Centrality based information blocking and influence minimization in online social network," in *Proc. IEEE Int. Conf. Adv. Netw. Telecommun. Syst. (ANTS)*, Dec. 2017, pp. 1–6.
[204] Q. Yao, R. Shi, C. Zhou, P. Wang, and L. Guo, "Topic-aware social influence minimization," in *Proc. 24th Int. Conf. World Wide Web*. New York, NY, USA: ACM, May 2015, pp. 139–140.
[205] C. Luo, K. Cui, X. Zheng, and D. Zeng, "Time critical disinformation influence minimization in online social networks," in *Proc. IEEE Joint Intell. Secur. Informat. Conf.*, Sep. 2014, pp. 68–74.
[206] J. A. Czepiel, "Word-of-mouth processes in the diffusion of a major technological innovation," *J. Marketing Res.*, vol. 11, no. 2, pp. 172–180, May 1974.
[207] T. N. Dinh, H. Zhang, D. T. Nguyen, and M. T. Thai, "Cost-effective viral marketing for time-critical campaigns in large-scale social networks," *IEEE/ACM Trans. Netw.*, vol. 22, no. 6, pp. 2001–2011, Dec. 2014.
[208] B.-S. Yan, F.-J. Jing, Y. Yang, and X.-D. Wang, "Network centrality in a virtual brand community: Exploring an antecedent and some consequences," *Social Behav. Personality, Int. J.*, vol. 42, no. 4, pp. 571–581, May 2014.
[209] S. Shao and C. Li, "Based on the social network evaluation model of short-term interaction with followers micro-blogging marketing," in *Proc. 8th Int. Workshop Semantic Social Media Adaptation Personalization*, Dec. 2013, pp. 8–13.
[210] C. Salavati, A. Abdollahpouri, and Z. Manbari, "Ranking nodes in complex networks based on local structure and improving closeness centrality," *Neurocomputing*, vol. 336, pp. 36–45, Apr. 2019.

[211] A. G. Nikolaev, R. Razib, and A. Kucheriya, "On efficient use of entropy centrality for social network analysis and community detection," *Social Netw.*, vol. 40, pp. 154–162, Jan. 2015.
[212] Y. Jiang, C. Jia, and J. Yu, "An efficient community detection method based on rank centrality," *Phys. A, Stat. Mech. Appl.*, vol. 392, no. 9, pp. 2182–2194, May 2013.
[213] R. M. Christley, G. L. Pinchbeck, R. G. Bowers, D. Clancy, N. P. French, R. Bennett, and J. Turner, "Infection in social networks: Using network analysis to identify high-risk individuals," *Amer. J. Epidemiol.*, vol. 162, no. 10, pp. 1024–1031, Nov. 2005.
[214] A. Dekker, "Network centrality and super-spreaders in infectious disease epidemiology," Tech. Rep., Dec. 2013.
[215] D. C. Bell, J. S. Atkinson, and J. W. Carlson, "Centrality measures for disease transmission networks," *Social Netw.*, vol. 21, no. 1, pp. 1–21, Jan. 1999.
[216] J. M. Gomez, C. L. Nunn, and M. Verdú, "Centrality in primate-parasite networks reveals the potential for the transmission of emerging infectious diseases to humans," *Proc. Nat. Acad. Sci. USA*, vol. 110, no. 19, pp. 7738–7741, May 2013.
[217] S. Gao, Y. Wang, Y. Gao, and Y. Liu, "Understanding urban traffic-flow characteristics: A rethinking of betweenness centrality," *Environ. Planning B, Planning Design*, vol. 40, no. 1, pp. 135–153, Feb. 2013.
[218] S. Porta, P. Crucitti, and V. Latora, "Multiple centrality assessment in Parma: A network analysis of paths and open spaces," *Urban Design Int.*, vol. 13, no. 1, pp. 41–50, Mar. 2008.
[219] R. Guimera, S. Mossa, A. Turtschi, and L. A. N. Amaral, "The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles," *Proc. Nat. Acad. Sci. USA*, vol. 102, no. 22, pp. 7794–7799, May 2005.
[220] Z. Li, M. Xu, and Y. Shi, "Centrality in global shipping network basing on worldwide shipping areas," *GeoJournal*, vol. 80, no. 1, pp. 47–60, Feb. 2015.
[221] H. Jeong, S. P. Mason, A.-L. Barabási, and Z. N. Oltvai, "Lethality and centrality in protein networks," *Nature*, vol. 411, no. 6833, pp. 41–42, May 2001.
[222] J.-H. Cho, S. Xu, P. M. Hurley, M. Mackay, T. Benjamin, and M. Beaumont, "STRAM: Measuring the trustworthiness of computer-based systems," *ACM Comput. Surveys*, vol. 51, no. 6, pp. 1–47, Feb. 2019, doi: 10.1145/3277666.



ZELIN WAN received the B.S. degree in computer science from The University of Arizona, Tucson, AZ, USA, in 2019. He is currently pursuing the Ph.D. degree with the Department Computer Science, Virginia Tech, Blacksburg, VA, USA. His research interests include game theoretic and machine learning-based cybersecurity and network science.

YASH MAHAJAN received the B.Tech. degree in computer science with specialization in bioinformatics from the Vellore Institute of Technology (VIT), Vellore, India, in 2019. He is currently pursuing the M.S. degree in computer science and applications with Virginia Tech, Blacksburg, VA, USA. His research interests include security and privacy in online social networks.

BEOM WOO KANG received the B.S. degree in electronic engineering from Hanyang University, in 2020, where he is currently pursuing the M.S. degree with the Department of Computer Science. He worked as a Research Assistant with the Trustworthy Cyberspace Laboratory, supervised by Dr. Jin-Hee Cho, in 2019. His current research interests include machine learning and computer systems.

TERRENCE J. MOORE (Member, IEEE) received the B.S. and M.A. degrees in mathematics from American University, in 1998 and 2000, respectively, and the Ph.D. degree in mathematics from the University of Maryland, College Park, MD, USA, in 2010. He is currently a Researcher with the Network Science Division, U.S. Army Research Laboratory. His research interests include sampling theory, constrained statistical inference, stochastic optimization, network security, geometric and topological applications in networks, and network science.

JIN-HEE CHO (Senior Member, IEEE) received the M.S. and Ph.D. degrees in computer science from Virginia Tech, in 2004 and 2008, respectively. Since 2009, she has been working as a Computer Scientist with the U.S. Army Research Laboratory (USARL), Adelphi, MD, USA. Since August 2018, she has been an Associate Professor with the Department of Computer Science, Virginia Tech, Blacksburg, VA, USA, and the Director of the Trustworthy Cyberspace Laboratory. She has published over 150 peer-reviewed technical articles in leading journals and conferences in the areas of trust management, cybersecurity, metrics and measurements, network performance analysis, resource allocation, agent-based modeling, uncertainty reasoning and analysis, information fusion/credibility, and social network analysis. She is a member of the ACM. She received the best paper awards in IEEE TrustCom'2009, BRIMS'2013, IEEE GLOBECOM'2017, 2017 ARL's Publication Award, and IEEE CogSima 2018. She is a winner of the 2015 IEEE Communications Society William R. Bennett Prize in the Field of Communications Networking. She was selected for the 2013 Presidential Early Career Award for Scientists and Engineers (PECASE), in 2016, which is the highest honor bestowed by the U.S. Government on outstanding scientists and engineers in the early stages of their independent research careers. She is also serving on the Editorial Board as an Associate Editor for *The Computer Journal* and IEEE TRANSACTIONS ON NETWORK AND SERVICE MANAGEMENT.