

RESEARCH

Open Access



# Node-weighted centrality: a new way of centrality hybridization

Anuj Singh<sup>1</sup>, Rishi Ranjan Singh<sup>1\*</sup> and S. R. S. Iyengar<sup>2</sup>

\*Correspondence:

rishi@iitbhilai.ac.in

<sup>1</sup> Department of Electrical Engineering and Computer Science, Indian Institute of Technology Bhilai, Raipur, Chhattisgarh, India

Full list of author information is available at the end of the article

This is an expanded and extended version of the results appeared in *CSoNet* 2019 [1].

## Abstract

Centrality measures have been proved to be a salient computational science tool for analyzing networks in the last two to three decades aiding many problems in the domain of computer science, economics, physics, and sociology. With increasing complexity and vividness in the network analysis problems, there is a need to modify the existing traditional centrality measures. Weighted centrality measures usually consider weights on the edges and assume the weights on the nodes to be uniform. One of the main reasons for this assumption is the hardness and challenges in mapping the nodes to their corresponding weights. In this paper, we propose a way to overcome this kind of limitation by hybridization of the traditional centrality measures. The hybridization is done by taking one of the centrality measures as a mapping function to generate weights on the nodes and then using the node weights in other centrality measures for better complex ranking.

**Keywords:** Complex network analysis, Centrality measures, Weighted networks, Hybrid centrality

## Introduction

Centrality measures are an important tool in social and complex network analysis to quantify the eminence of nodes. These measures remain invariant under isomorphic transformation of a network [2]. By definition, a centrality measure is a quantification of the structural importance of a node based on its location, connectivity, or any other structural property. These have been used not only by the network scientists, but also by biologists, sociologists, physicists, psychologists, and economists over time. Several measures are coined in literature. The most popular centrality measures for network analysis (traditional measures) are degree, closeness, betweenness and eigenvector centrality. Readers are referred to the books by Jackson [3] and Brandes and Erlebach [2] for a detailed survey on the centrality indices and their applications. There exist several other measures which either extend or generalize these traditional measures or limit them to a restricted application. Moreover, various variants of these centrality measures have been proposed which consider a set of nodes and compute its collective centrality, called group-centrality [4]. Yet another direction is to combine various centrality measures to achieve better results for answering more complex problems. Such measures

are termed as hybrid-centralities and few of them are summarized in section on related works.

In the last two decades, a major portion of the interdisciplinary work evolved just applying these measures to extract information from underlying network data. Each of the proposed centrality measures ranks nodes in a network based on a specific structural attribute, making them application specific. Due to this reason, choosing an appropriate centrality for a given application has been a key issue in network analysis. A major portion of the research work done in this direction is concerned with selecting the best of the available measures for a particular application or defining a new measure that outperforms the existing ones.

Most of the centrality measures proposed in literature were first defined for unweighted graphs, i.e., all of the nodes and all the edges were assumed homogeneous in the beginning of centrality computation. We refer these measures as unweighted centrality measures. After realizing the existence and understanding the necessity and importance of weights on the edges, centrality measures for unweighted graphs were extended to edge-weighted centrality measures. These measures take weights on the edges into consideration for ranking the nodes while analyzing the networks, but still assuming equal weights on the nodes. A substantial part of the present-day research in the analysis of weighted networks considers only edge weights to determine the topological significance of nodes [5].

We call a network with nonuniform weights on both: edges and nodes, as a fully weighted networks. Several such networks surround us. The weights on the nodes in fully weighted networks can be understood as some sort of mapping of the characteristics or attributes of the nodes to some real value. At times, these weights can also be dependent on the structure around the nodes. Let us understand the possibility of existence of weights on the nodes with the help of some popular networks around us. In these networks weights on the edges have already been discussed above. Therefore, here, we only discuss regarding the existence of nonuniform weights on the nodes.

1. Friendship/social networks: in this type of networks, nodes represent persons and edges represent the friendship relationship between the considered set of persons. Here, weights on the nodes can be understood as a mapping of wealth, power, education level, or some other attribute of persons. It is notable that existence of two persons with identical attributes is highly unlikely and therefore all person's attributes can be mapped to different real values based on the application specific mapping.
2. Public transit networks: road networks, train networks, metro rail networks, airline networks are some of the major public transit networks. In these types of networks, nodes represent locations (place) and edges represent direct connected traveling medium between the locations. In such a network, weights on the nodes can be understood as a mapping of the population, frequency of commuters, development status, popularity for tourism, etc., of the location represented by that node.
3. Communication networks: in communication networks, nodes are represented by the communicating devices and links represent the direct (wired or wireless) connection between these devices. Weights on the nodes in such networks can be understood as the role, cost, location of these devices.

4. Co-authorship network: in such type of networks, nodes represent authors and edges represent the co-authorship relation, i.e., at least one article was written by the authors across each edge as co-authors. In such a network, weights on nodes can represent the popularity, h-index, citations, awards, designation of the authors, etc.

Similar to the networks stated above, several other kinds of networks also fall into the category of fully weighted networks. It should be noted that in any network, all nodes (actor/devices/places/person) with same characteristics is highly unlikely. One basic reason to ignore the weights on nodes is the hardness and complexity in figuring out the non-trivial mapping from attributes of nodes to real values. Another reason is to avoid the complexity of the analysis process, but sometimes the ignorance of node weights might lead to a wrong analysis of the given network. We show through toy examples that the unweighted and edge-weighted centrality measures may exhibit limitation in correctly analyzing the fully weighted networks. Therefore, it becomes essential to also consider weights on nodes in the process for a better analysis of fully weighted networks. Due to this reason, there was a need to upgrade the previously defined unweighted and edge-weighted centrality measures so that weights at the nodes can also be taken into account for the analysis.

Therefore, while considering the weights on the nodes, there can be two possible extensions of the unweighted and edge-weighted centrality measures:

- Node-weighted centrality measures: consider only the node weights for the analysis while taking all edge weights as one.
- Fully weighted centrality measures: consider both types of weights, edge weights and node weights for the analysis.

In most of the studies done so far on fully weighted networks, while considering weighted edges, weights on the nodes were completely ignored. Meanwhile, only a little work has been accomplished while considering the weights on nodes in the real-world networks [6–9]. The main contribution of this paper is to motivate the analysis of networks while considering the weights on the nodes. In order to overcome the challenges in figuring out the weights on the nodes, we propose to use appropriate centrality measures to generate the weights and then further use these weights in the node-weighted centrality measures to analyze a given network. We give two applications based on this principle of hybridization.

In the next section, we provide the basic notations and provide definitions of traditional unweighted centrality measures. Two toy examples pointing out and motivating the necessity of considering weights on the nodes is discussed after that. Next, node-weighted centrality measures and related works are summarized. Afterwards, new hybridization of centrality measures are introduced for solving two complex computational problems in networks based on the definitions of node-weighted centrality measures. The experimental comparison between the newly proposed hybrid centrality measures and traditional centrality measures on real-world networks is comprised in the respective sections on problems. Finally, we conclude and discuss the prospective future directions.

It is an expanded and extended version of the results that appeared in CSoNet 2019 [1]. The conference version titled “Hybrid Centrality Measures for Service Coverage Problem” discussed a hybridization of centrality measures to solve Service Coverage problem based on the formulation for node-weighted centrality measures. In this version, we have extended the concept of this type of hybridization to solve another problem called spread maximization problem. Due to this, the structure of the paper has been revised and this version is focused on motivating hybridization of centrality measures based on the formulation of node-weighted centrality measures. Application of the proposed measures to more than one problem in this version exhibits the potential for this kind of hybridization and motivates for several open directions in this area.

### Unweighted centrality measures

This section first defines the basic notations used through out the paper. Then, we briefly summarize traditional centrality measures; degree, closeness, betweenness, and eigenvector. We also describe in brief harmonic centrality that is highly correlated to closeness centrality.

Let  $G = (V, E)$  be an undirected unweighted network, where  $V$  is the set of nodes and  $E$  is the set of links. Let the number of nodes in the network, i.e., the cardinality of set  $V$  be  $|V| = n$  and the number of links, i.e., the cardinality of set  $E$  be  $|E| = m$ . Let  $A$  be the adjacency matrix of size  $n \times n$  in network  $G$ , where entry  $a_{ij}$  denotes whether there is a link between node  $i$  and node  $j$ .  $a_{ij} = 1$  if there exists a link between node  $i$  and node  $j$ , otherwise  $a_{ij} = 0$ . Let  $d_{ij}$  be the geodesic distance, length of the shortest path, between node  $i$  and node  $j$ . Next, we discuss in brief some of the widely used centrality measures.

- Degree: Freeman’s [10] degree centrality considers that a node’s importance is proportional to its degree, number of links connected (starting/ending) to that node. Mathematically, degree centrality of a node  $u$ ,

$$DC(u) = \sum_{v \in V \setminus \{u\}} a_{uv}.$$

Degree centrality can be normalized by dividing the above expression with  $n - 1$ . In a social network, degree centrality of a node represents that node’s popularity. A higher degree node has many followers/friends which shows the strength of the node. Lower degree nodes are the actors who are nearly isolated from the population of the network and are least popular.

- Closeness: Freeman’s [10] closeness centrality considers a node’s importance to be inversely proportional to the sum of its distance to other nodes. Mathematically, closeness centrality of a node  $u$  can be represented as

$$CC(u) = \frac{1}{\sum_{v \in V \setminus \{u\}} d_{uv}}.$$

Closeness centrality can be normalized by multiplying the above expression with  $n - 1$ . The concept of closeness centrality was first given by Freeman [10] for social networks, but the concept has existed for a long time as status of a node [11]. Closeness centrality of a node in a network represents the node’s average distance, i.e., the

expectation of how much time, a piece of information that started flowing from any node in the network, will take to reach that node or vice-versa. A higher closeness central node gets updated very early when some information is spreading and nodes with low closeness centrality have to wait longer for getting updated with the flowing information.

- Harmonic: harmonic centrality [5, 12] considers that a node's importance in a network is proportional to the sum of inverse of its distances from other nodes. Mathematically, harmonic centrality of a node  $u$  is

$$HC(u) = \sum_{v \in V \setminus \{u\}} \frac{1}{d_{uv}}.$$

Freeman's [10] closeness centrality was not applicable on disconnected networks. Harmonic centrality is highly correlated to closeness centrality [12] and works well on disconnected and directed networks.

- Decay: this centrality works on same principle as harmonic centrality but at place of penalizing the contribution of nodes linearly, it does exponentially [3]. Mathematically, decay centrality of a node  $u$  is,

$$DKC(u) = \sum_{v \in V \setminus \{u\}} \delta^{d_{uv}},$$

where  $\delta$  lies between 0 and 1.  $\delta$  is called decay parameter.

- Betweenness: Freeman's [10] betweenness centrality considers that a node's importance is proportional to the number of times that node occurs in the shortest paths between all possible pairs of nodes in a network. Mathematically, betweenness centrality of a node  $u$ ,

$$BC(u) = \sum_{s,t,u \in V: | \{s,t,v\} | = 3} \frac{\sigma_{st}(u)}{\sigma_{st}},$$

where  $\sigma_{st}$  is the total number of shortest paths from node  $s$  to node  $t$  and  $\sigma_{st}(u)$  is the total number of shortest paths from node  $s$  to node  $t$  passing through node  $u$ . Betweenness centrality can be normalized by dividing with  $\frac{(n-1)(n-2)}{2}$ . The concept of betweenness centrality was first proposed by Anthonisse [13] and Freeman [14] independently. Betweenness centrality of a node in a network represents the node's brokerage power, i.e., control over the flow passing through the network with the assumption that information is flowing through shortest paths. A higher betweenness central node controls a major fraction of flow passing through the network while a low betweenness central node has nearly no such control.

- Eigenvector: Bonacich's [15] eigenvector centrality considers that a node's importance in a network is proportional to the sum of the importance of neighboring nodes in that network. Mathematically, eigenvector centrality of a node  $u$  can be written as,

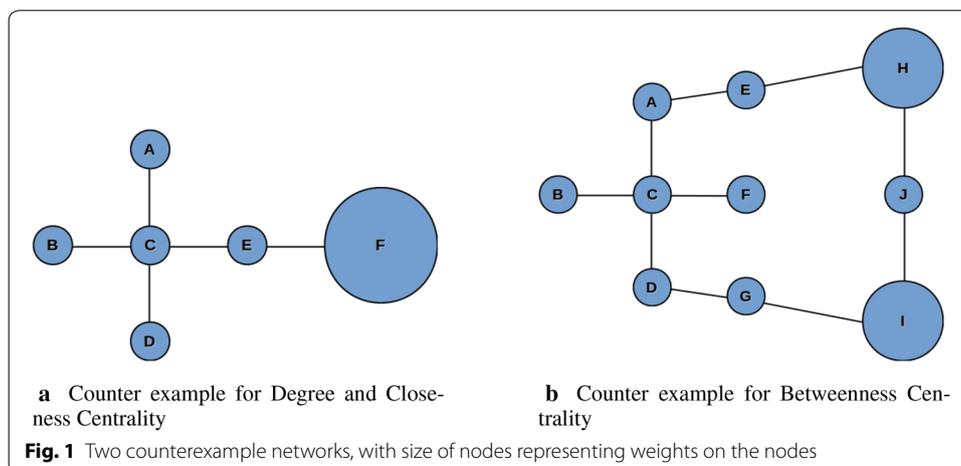
$$EC(u) = \sum_{v \in V \setminus \{u\}} (a_{uv} \cdot EC(v)).$$

Eigenvector centrality of a node in a network represents the power of a node’s neighbors in the network. A node with high eigenvector centrality indicates the direct connection of this node to important nodes in the network and vice-versa.

**Toy examples**

Consider the example networks given in Fig. 1. Let us assume that the first network in Fig. 1 is a road network connecting the cities of a state where the size of nodes represent the weights (in this case population) on the nodes. City F is highly populated while Population in other cities is significantly low and several times lesser than city F. Average-sum facility location problem questions for a node in a given network that is at the minimum average distance to all other nodes in the network. Now, if we are attempting to solve the average-sum facility location problem to install a facility for the whole population of the state, the solution is the most closeness central node assuming equal population in each city. By the definition of closeness centrality, the answer is city C, but this city is not really suitable in the reality for the whole population. The answer is city F where most of the population of the state already resides. Now, suppose if we are tackling a problem where the goal is to find a city from which maximum population is exactly at one hop distance, degree centrality seems to solve this problem in unweighted networks but here it fails. Degree centrality ranks city C as the most central but because of nonuniform distribution of population, city E is the correct answer.

Next, let us assume that the second example network given in Fig. 1 is a communication network where the size of nodes denotes weights (in this case we can assume importance/vitality) of the nodes. We assume that communication happens through shortest paths and the importance of a communication is a function of the importance of nodes, between which communication is happening. Here, we assume that the function computes the multiplication between the importance of the communicating nodes. Now, the goal is to compute the node which controls the most important communications in the network. Applying betweenness centrality, which is a tool directly used for this type of problems, answers node C neglecting the importance attribute. In the reality, node J is the answer because it controls the communication between the most important nodes in the given network. We note above that the existing centrality measures



neglect node weights and due to this, are prone to analyze incorrectly the fully weighted networks. Therefore, there is a need to upgrade the current definitions of unweighted or edge-weighted centrality measure.

### Node-weighted centrality

This section summarizes centrality measures that take into consideration the weights on the nodes while giving equal priority to the edges in a given network. These fall in the category of node-weighted centrality measures. We only mention these here to avoid the complexity of including weights on both edges and nodes. Once, these measures are understood, we recommend the readers to combine these measures with the edge-weighted centrality measures given in [5] to derive the definition for fully weighted centrality measures. Recall  $G = \{V, E\}$  as the undirected unweighted graphs. We add an extra element, weights on the nodes defined as a function  $W : V \rightarrow R$ , where  $R$  is set of real numbers. Let  $W_x$  be the weight given at node  $x$ . We do not directly use  $W_x$  in the definitions, but at place of it, we use a function  $f$  of  $W_x$  (or a function of  $W_x$  and  $W_y$  depending on the number of parameters) without losing the generality. This gives us the flexibility to tune the function of weights on the nodes according to our need. Here, in this paper we take  $f(W_x) = W_x$  for the simplicity. To normalize the new centrality measures, we divide by the maximum possible value that any node can score in the numerator of below-given formulas. We start with degree centrality.

- Node-weighted degree centrality: in [6], weights on nodes are considered and the definition of degree centrality is modified to accommodate the node weights. Abbasi and Hossain [6] considered centrality scores as weights on the nodes. Following it, node-weighted centrality of a node  $u$  is calculated as:

$$\frac{\sum_{v \in V \setminus \{u\}} (f(W_v) \cdot a_{u,v})}{\sum_{x \in V} f(W_x)}.$$

This measure assigns higher importance to those nodes which are in the immediate neighborhood of highly weighted nodes. Next two measures extend the consideration to all the nodes to compute the eminence of a node.

- Node-weighted closeness/harmonic centrality: to target the wider applicability, we define node-weighted harmonic centrality (which also can be used in the case of closeness centrality computation as both are highly correlated). Node-weighted harmonic centrality of a node  $u$  in a network is defined as:

$$\frac{f(W_u) + \sum_{v \in V \setminus \{u\}} \frac{f(W_v)}{d_{u,v} + 1}}{\sum_{x \in V} f(W_x)}.$$

This measure depends on two factors: weight of the node  $u$  under consideration and the effective weights of other nodes corresponding to their distances from node  $u$ . It assigns a higher value to the nodes that are of high weights and closer to the nodes with high weights. We refer this measure as harmonically attenuated node-weighted centrality measure.

- Node weighted decay: weighted decay of a node  $u$  in a network as defined as

$$\frac{f(W_u) + \sum_{v \in V \setminus \{u\}} (\delta^{d_{u,v}} \cdot f(W_v))}{\sum_{x \in V} f(W_x)},$$

where  $\delta$  lies between 0 and 1. Here also the computation of importance depends on the same two factors as in NWCC. But, the contribution of weights of other nodes decays exponentially with distance. Weighted decay assigns a higher value to the nodes that are of high weights and very close to the nodes with high weights. We refer this measure as exponentially attenuated node weighted centrality measure.

- Node-weighted betweenness centrality: in [16], a factor denoting the importance of communication between two pairs is multiplied in the core formula while computing the betweenness centrality. The node-weighted betweenness centrality of a node  $u$  was defined as

$$\frac{\sum_{s,t,u \in V: | \{s,t,v\} | = 3} f(W_s, W_t) \cdot \frac{\sigma_{st}(u)}{\sigma_{st}}}{\sum_{x,y,u \in V: | \{x,y,v\} | = 3} f(W_x, W_y)},$$

where  $f(W_x, W_y)$  can be assumed to map the weights given on node  $x$  and  $y$  to a real value denoting the importance of flow happening between  $x$  and  $y$ .

Eigenvector centrality is a measure where it is still open how to include the effect of node weights. Even if we start the eigenvector centrality computation with a vector comprising the weights on the nodes, at the time when the convergence occurs, we arrive at the same eigenvector as the solution every time. It is because the computation of this particular centrality is dependent only on the adjacency matrix and eigenvector is the property of this matrix only. One way around here is to follow the idea given in [9], where they multiply the  $\beta$ th power of the weight of each node to its unweighted centrality to compute node-weighted centrality. Here,  $\beta$  takes a value from the interval  $[-1, 1]$ .

In this paper, we do not explore about these measure in the experimental section. We plan to do a thorough application oriented study of the above measures in near future. We present the definition for the sake of completeness of major traditional centrality measures.

### Related work

Centrality measures are the tools to find application specific importance of a node. Unweighted Centrality measures mentioned earlier, are the most widely used measures but for complex problems and applications, these measures are inefficient. In that case, a combination of these centrality measures produces better analysis than using them individually. In a recent study [6], authors have proposed a new set of hybrid centralities; degree-degree, degree closeness, and degree betweenness. They noticed that the newly defined centrality measures are different than the traditional centrality measures. On real-world co-authorship network, they found that all the newly defined hybrid centrality measures are significantly correlated to authors' performance (sum of citations of authors' h-index). A similar study [7] on weighted collaboration networks was done and three new sets of collaborative centrality measures were proposed. The traditional collaborative centrality measures for an author node (h-index, a-index, g-index) are used to propose new centrality measures. Newly defined set of measures were highly

correlated with the traditional performance measures of scholars (publication count, citation count, h-index, a-index, g-index). Zhang et al. [17] proposed a new hybrid centrality measure to measure a node's importance in satellite communication network. This measure is also based on the combination of closeness and betweenness centrality but the considered measure in their paper punishes the betweenness importance with a different factor. Qiao et al. [18] proposed a hybrid page scoring algorithm based on degree centrality, betweenness centrality, closeness centrality and the PageRank algorithm. Lee and Djauhari [19] proposed a hybrid centrality measure which was a linear combination of the degree, closeness, betweenness, and eigenvector centrality measures. The proposed measures were used to find the most significant or influential stocks. A hybrid centrality based on the linear combination of degree centrality and cohesion centrality was proposed was Qiu et al. [20] and further used for community detection by Li-Qing et al. [21]. Wang et al. [22] proposed a hybrid centrality measure based on the combination of degree, the degree of neighbors and betweenness. In another study by Buechel and Buskens [23], authors analyze a hybrid centrality model as a combination of extended closeness (a variation of closeness for disconnected networks), betweenness and degree measures. None of the above studies attempt to solve the service coverage problem.

### **The service coverage problem**

This section covers the first of two applications covered in this paper, the service coverage problem (SCP). We present two new hybrid measures to solve SCP a type of complex facility location problem. We define these measures for networks without weights on edges. These can be easily further extended to edge-weighted version of hybrid centrality measures following the ideas for edge-weighted centrality measures given in [5].

Flow networks are those networks where information, people, commodities, etc., move from one node to other while traversing on the links. Each node starts with a limited bandwidth of resources for convening the flow and these resources prone to collapse/degrade/reduce over time due to the continuous participation of the node in fostering the flow. Due to this, such networks require uninterrupted maintenance/replenishment service on a regular basis for the proper functioning of the network. Keeping this in mind, service facilities are installed at nodes that meet the service demand of other node from time to time.

After a service request is made by some node for more resources or reviving the node after a failure occurs, the service team is sent from the service center to meet the demand. The response-time to meet the service demand is defined as the time taken between the request for service and start of the service work on the demanding node. The response-time depends on the distance between the node requesting for a service and the node with the service stations. It is under the assumption that the service centers have sufficient resources to meet the demand of other nodes. The response-efficiency of a node is inversely correlated to the response-time, i.e., when the response-time is least, the node is said to be maximum response-efficient and most suitable for installing service stations. A node with a higher response-time possesses a smaller response-efficiency and is not appropriate choice for installing service stations.

Given an unweighted network, the objective of service coverage problem (SCP) is to find a best suited node to install service station facility such that the expected response-time is reduced. In another word, the goal is to find a node with highest expected response-efficiency, i.e., the service should reach faster at the nodes that request for service more often than the nodes with moderate or seldom demands. Therefore, service centers should be installed closer to the nodes with frequent demands.

Randomly choosing a node to install service facility is certainly not the best solution. It is because the randomly picked node may be very far from the nodes with higher demand. In such a case where a node with higher load fails, severe damage would have already been caused before the maintenance service reaches it. Thus, we require a measure to evaluate the importance of nodes for the candidacy to install stations covering service requests.

Betweenness centrality is used to predict the importance of nodes in a number of flow based networks, e.g., power grid networks, public-transit networks, gas line networks, and communication networks. The betweenness scores of the nodes in such networks have been considered as the load on the nodes in many literature. Several models for understanding and replicating the cascading phenomena have been proposed [24–26]. Few of these models observe that failure of a node with a high load may cause a severe cascading failure and hence result in the breakdown of the whole network system. After a node fails, the best way to reduce the damage is to recover the failed node as soon as possible. We cannot prevent a node from failure but we can definitely put a maintenance mechanism on some of the nodes in the network.

### Formulation

In the service coverage problem in flow networks, the load of the nodes (when no other information is provided) can be assumed to be proportional to the betweenness centrality of the nodes. This is because a node with large flow through it, is expected to degrade faster than other nodes and if such a node shuts down, a large amount of flow will be affected. Thus we take the probability of node  $j$  requesting for a maintenance service over a fixed time-interval as

$$\Pr[\text{Node } j \text{ request for service}] = \frac{BC(j)}{\sum_{k \in V} BC(k)},$$

where  $BC(j)$  is the betweenness centrality of node  $j$ . Let  $X_{ij}$ , be the response-efficiency of a node  $i$  to meet service demand from node  $j$  if the service station is going to be installed on node  $i$ . Certainly,  $X_{ij}$  depends on the distance between the node requesting for a service (node  $j$ ) and the node with the service stations (node  $i$ ). Harmonic decay can be considered in applications where the nodes are more robust and the service requirement is not urgent while exponential decay might be a more appropriate simplified model in applications like real-time systems where service requests need to be met on an urgent basis.

If the response-efficiency decays harmonically in a given application, i.e., the node is maximum efficient (efficiency value = 1) for itself while it is half efficient (efficiency value = 0.5) for each of its neighbors and so on, then we can formulate for the value of response-efficiency  $X_{ij} = \frac{1}{d_{i,j}+1}$ . Recall,  $d_{u,v}$  denotes the distance, i.e., length of the

shortest path from node  $u$  to node  $v$  in the given network. Let  $\chi_i$  be the total response-efficiency of a node  $i$  if the service station is going to be installed on node  $i$ . Then,

$$\chi_i = \sum_{j \in V} X_{ij} \cdot \Pr[\text{Node } j \text{ request for service}].$$

The expected response-efficiency of node  $i$  ( $E[\chi_i]$ ) is computed by taking expectation over that node's response-efficiency to service all the nodes in network:

$$E[\chi_i] = \sum_{j \in V} \frac{1}{d_{i,j} + 1} \cdot \frac{BC(j)}{\sum_{k \in V} BC(k)}.$$

We define the following hybrid centrality measures based on the above formulation:

- Harmonically attenuated-betweenness centrality (HABC): this measure is based on the harmonic attenuation of importance with respect to the distance. The harmonically attenuated-betweenness centrality of a node  $u$  is defined as,

$$\text{HABC}(u) = \frac{BC(u) + \sum_{v \in V \setminus \{u\}} \frac{BC(v)}{d_{u,v} + 1}}{\sum_{w \in V} BC(w)},$$

where  $BC(u)$  is the unweighted betweenness centrality of node  $u$ . This hybrid measure assigns a value that is calculated based on an averaging function of distances and betweenness centrality. This measure assigns a higher value to the nodes that are of high betweenness and closer to other high betweenness central nodes. This measure solves the SCP problem in flow networks where the response-efficiency decays harmonically.

In few complex systems, response-time plays very crucial roles in the functionality of networks and the response-efficiency decays exponentially. In such networks, the response-efficiency decays faster, i.e., the node is maximum efficient (efficiency value=1) for itself while it decays by a multiplicative factor  $\alpha$  (efficiency value =  $\alpha$ ) for each of it's neighbors and so on. Here,  $\alpha$  is a severity factor that lies in interval[0,1]. Then we can formulate for the value of response-efficiency  $X_{ij} = \alpha^{d_{ij}}$ . Similar to the previous analysis, the expected efficiency of a node  $i$  to be a service station, when efficiency decays exponentially by a factor  $\alpha$  is  $E[\chi_i] = \sum_{j \in V} \alpha^{d_{ij}} \cdot \frac{BC(j)}{\sum_{k \in V} BC(k)}$ . We define the following hybrid centrality measures based on the above formulation:

- Exponentially attenuated-betweenness centrality (EABC): this measure is based on an exponential attenuation in the power of distance. The exponentially attenuated-betweenness centrality of a node  $u$  is defined as,

$$\text{EABC}(u) = \frac{BC(u) + \sum_{v \in V \setminus \{u\}} \alpha^{d_{u,v}} \cdot BC(v)}{\sum_{w \in V} BC(w)},$$

where  $BC(u)$  is the betweenness centrality of node  $i$ .  $\alpha$  is the attenuation factor that is used in the exponential function to the power of distance. It is used to sharply punish the service time. This hybrid measure assigns a value that is calculated based on betweenness centrality of nodes and an exponential function in the power of distances. This measure assigns a higher value to the nodes that are of high betweenness

and very close to other high betweenness central nodes. This centrality measure is a sharp attenuation variant of harmonically attenuated-degree centrality. This measure solves the service coverage problem when the service is required on a very urgent basis.

Let  $m$  be the number of links and  $n$  be the number of nodes in a given network. The over all time to compute proposed hybrid centrality measure is  $O(mn)$  in unweighted and  $O(mn + n^2 \log n)$  in weighted graphs. It is due to the time complexity for computing betweenness centrality [27]. Efficient algorithms for updating and estimating betweenness centrality measures in dynamic and large-size networks are discussed in [28–30].

### Experimental results for service coverage problem

In this section, we discuss the simulation results on various real-world networks. First, we discuss the experimental setup. Then, we mention all the data set used for experimentation. Next, we provide a comparison of traditional centrality measures: degree (DC), closeness (CC), betweenness (BC) and, eigenvector (EC) and the proposed hybrid centrality measure [HABC, EABC ( $\alpha = 0.5$ ), EABC ( $\alpha = 0.75$ )] in the considered networks. The experiments are performed on a Linux system with 8 GB RAM and Intel Core i5-8250U CPU at 1.6 GHz. Implementation is done in Python 3.6.7 with the help of Networkx library.

*Considered real-world data sets* the proposed solution to solve service coverage problem discussed in this paper hybridizes betweenness centrality within closeness/harmonic and decay centrality. Betweenness centrality is first used to compute the load on each node, therefore, we have selected 21 real-world flow networks. We provide a brief summary of these networks in Table 1 and [31–39] can be referred for a detailed description of the networks. We have considered various types of transport networks, energy networks, internet peer-to-peer network, etc. The columns of Table 1 consist of names of the network instances, the number of nodes ( $n$ ), the number of edges ( $m$ ), the average degree of the networks (Avg. deg.), density of the network, average clustering coefficient ( $\hat{C}$ ), degree assortativity, the size of the maximum clique and the network type, respectively.

*Simulation* we have conducted simulations for evaluating the performance of various traditional centrality measures and the hybrid measures proposed in this paper which are summarized in Table 2. The performance of a centrality measure is evaluated by computing the expected response-time in terms of the average distance of the service requesting nodes from the top central node as per this centrality measure. We have emphasized in bold, the best expected response time in Table 2 for each network instances given in Table 1, when a service maintenance center has been installed in a node picked according to the considered centrality measures.

Betweenness centrality has been used as one of the best measure to map loads on nodes in flow networks [24]. A node having higher load will be more frequent in requesting for services. Therefore, we consider the probability of a node requesting for services proportional to the betweenness centrality of that node. The expected response-time is computed over  $\lceil n/10 \rceil$  service requests where  $n$  is the number of nodes in respective real-world networks. The first column in the table contains label of the considered real-world network instances. The next columns lists the expected response-time of the

traditional centrality measures (BC, CC, DC, EC) and the proposed hybrid measures (HABC, EABC( $\alpha = 0.5$ ), and EABC( $\alpha = 0.75$ )).

It is evident from Table 2 that no traditional measure can consistently find an ideal service center. While at least one of the proposed hybrid centrality measures are always able to minimize the average response-time for all considered networks. In some cases such as Madrid Metro, Paris Metro, Osaka Metro, Seoul Metro, and Tokyo Metro, the proposed measures even outperform all the traditional measures.

*Node ranks* here, we discuss the ranking result of nodes by various centrality measures. Table 3 contains the ranking list of the top three central nodes using the proposed hybrid measures and the traditional centrality measures on the considered real-world networks given in Table 1. Due to space constraints, only the data of top three central nodes have been presented in the paper. The first column in the table contains the label of the considered real-world network instances. The next columns in a group of three lists top 3 central nodes using the proposed hybrid measures (HABC, EABC( $\alpha = 0.5$ ), and EABC( $\alpha = 0.75$ )) and the traditional centrality measures (BC, CC, DC, EC).

Finding only the most central node might not be useful in the cases when the top central node does not allow it to be made as a service facility due to some constraints. In that case, finding the first top- $k$  potential centers are important. It is evident from the experimental results that the proposed measures are results of hybridizations between closeness centrality and betweenness centrality.

We have emphasized the top-ranking node for a few entries in the table and have written them in bold to exhibit the above phenomenon. For some networks, the top-ranking node as per HABC is the same as BC but not as CC and vice-versa. The ranking on two networks (Minnesota Road and Moscow Metro) also provide evidence that top-ranking nodes due to the proposed centrality measures are different from closeness and betweenness centrality. In addition to these two, another network (Openflights) shows that these proposed measures also rank nodes differently than each other.

It is evident that no single standard centrality measure is a perfect substitute for our proposed centrality measures. As our measures follow directly from the requirements of the Service Coverage Problem, it is evident that current centrality measures are not sufficient to solve the problem adequately.

We present the Spearman's rank correlation coefficient, Kendall's rank correlation coefficient and Fagin's intersection metric [40] to evaluate the correlation between traditional centrality measures and the proposed hybrid centrality measures in Tables 4, 5 and 6, respectively. For our purposes we compute the Fagin's rank intersection for top-1000 ranks in the network. In case the number of nodes is less than or equal to 1000, the rank intersection is computed for all nodes in the network. The first column in the tables contain the label of the considered real-world network instances. The next four columns comprise the rank correlation between HABC and the traditional centrality measures (BC, CC, DC, EC). Similarly the next four columns contain the rank correlation values between and EABC( $\alpha = 0.5$ ) and BC, CC, DC, EC, respectively. The last four columns are the rank correlation between EABC( $\alpha = 0.75$ ) and the traditional centrality measures.

The experimental results in Tables 4 and 5 for the Spearman's and Kendall's rank correlation coefficients make it evident that ranking by the proposed hybrid measures is different than traditional centrality measures. The average correlation coefficient between

the proposed measures and traditional measures, although is positive but very small. The standard deviation is larger than the average coefficient values for most of the computation of rank correlation. It is due to several negative correlation coefficient values. Based on the average rank correlation coefficient values, HABC is best correlated with CC and then BC among the traditional measures.  $EABC(\alpha = 0.75)$  also exhibits similar pattern as the decay rate is slower than  $EABC(\alpha = 0.5)$ .  $EABC(\alpha = 0.5)$  is best correlated with the CC and then EC among the traditional measures. The proposed measures are least correlated with DC. The high Fagin's intersection is due to the ranks being restricted to the top-1000. Therefore, the existing traditional measures do not provide the solution to the service coverage problem. The top-ranked nodes by the proposed hybrid centrality measures are more appropriate in this application.

### **Spread maximization in complex contagion**

In this section, we present the second application. Contagions are the phenomena of disease/behavior/trends/information/idea spreading across networks. Contagions can be classified as simple or complex. Simple contagions can be transmitted by a single infected individual to other. Complex contagions, however, require multiple exposures to infect an individual. Instead of simple contagions like diseases, there have been studies such as [41] and [42] which indicate that the spread of ideas, trends, behavior, influence and information in a social network can be more accurately modeled as complex contagions. In a recent study by [43] on real-world networks, it has been noted that influence maximization is not due to the nodes in higher cores. We consider complex contagions first in a linear threshold diffusion model setting [44, 45]. Linear threshold models are one of the classical diffusion models. The considered linear threshold model in this paper assumes that a fixed proportion of neighbors of an individual need to be infected in order to transmit infection to that individual. It is also referred as relative threshold model. There are other models that considered deterministic thresholds yet distinct for each individual. Another deterministic threshold model fixes the threshold to be a constant number and does not depend on the neighborhood size. In stochastic setting, uniformly random thresholds from an interval are considered for individuals.

We also consider two stochastic models of diffusion, independent cascade and stochastic linear threshold. In the independent cascade model every edge in the network is associated with a random value which represents the probability of diffusion across the edge. The stochastic linear threshold model is a linear threshold model in a stochastic setting as explained earlier.

The total number of people infected can depend upon where the contagion starts, threshold on individual node, etc. There can be multiple sources in a network which can start a complex contagion. This is especially true in the case for complex contagions, which have a hard time jumping across communities [46]. Hence identification of an ideal origin of trend/behavior/idea, etc., for a contagion can be helpful if we want the contagion to spread to maximum number of people. This has real-world benefits, such as in social marketing campaigns.

In this section, we consider the problem of finding a seed node that could spread a behavior/advertisement/idea/trend or influence to maximum number of nodes. It is referred as spread maximization problem. In complex contagion, a single node

might not have the strength to even start the propagation of infection, therefore, it is assumed that the neighborhood of the seed node is also infected with the seed node at time step zero.

### Formulation

In this section, we formulate hybrid centrality measures for the above stated problem of spread maximization in complex contagions. When choosing a source for infection, several factors come to mind. Out of which, the most important is that the node must have high enough degree to expose multiple individuals with the infection. It also helps if the node is closer to other nodes with high degree to spread it further. Hence to measure the potential of a node in acting as a top spreader, we use its degree as its node weight. We present the following two hybrid measures to identify ideal sources for the spread of complex contagions in the relative threshold model, leveraging the hybridization similar to the one used in previous section.

- Harmonically attenuated degree centrality (HADC): in this measure, we attenuate the node weights harmonically with respect to distance. Formally, for a node  $u$ , the  $HADC(u)$  can be expressed as

$$HADC(u) = \frac{DC(u) + \sum_{v \in V \setminus \{u\}} \frac{DC(v)}{d_{u,v} + 1}}{2m},$$

where  $DC(v)$  is the degree centrality of node  $v$ ,  $m$  is the number of links in the graph.

- Exponentially attenuated degree centrality (EADC): in this measure, we attenuate the node weights exponentially with respect to distance. Formally, for a node  $u$ , the  $EADC(u)$  can be expressed as,

$$EADC(u) = \frac{DC(u) + \sum_{v \in V \setminus \{u\}} \alpha^{d_{u,v}} \cdot DC(v)}{2m},$$

where  $\alpha$  is a parameter that lies between 0 and 1,  $m$  is the number of links in the graph,  $DC(v)$  is the degree centrality of node  $v$ .

Let  $m$  be the number of links and  $n$  be the number of nodes in a given network. The over all time to compute proposed hybrid centrality measure is  $O(mn)$  in unweighted and  $O(nm + n^2 \log n)$  in weighted graphs. It is due to the time complexity for all pair shortest paths. In real-world, each node has its own threshold which may be deterministic or stochastic in nature and varies from individual to individual. Finding best spreader node is possible in  $O(mn)$  time if the threshold values at all nodes are fixed to some constant value or proportion. Yet, for a different value of threshold, the best spreader may change and requires re-computation. The proposed measures do not require the knowledge of threshold value to be known beforehand. From the simulation results in the next section, it is evident that these measures have consistently performed the best in most cases for different values of threshold.

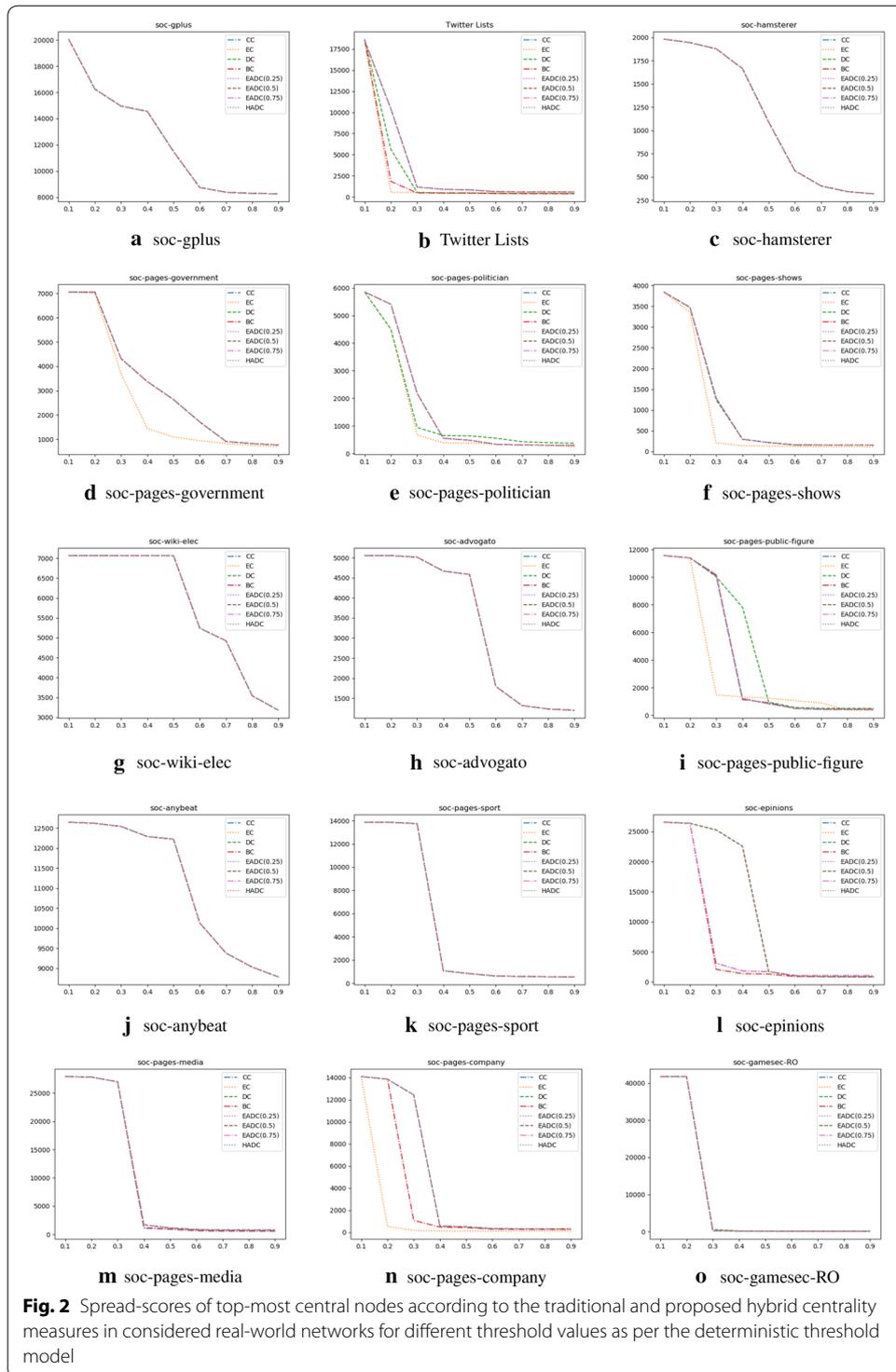
### Experimental results for spread maximization problem

In this section, we discuss the simulation results for for spread maximization problem on various real-world networks. First, we discuss the experimental setup. Then, we mention all the data set used for experimentation. Next, we provide a comparison of traditional centrality measures: degree (DC), closeness (CC), betweenness (BC) and, eigenvector (EC) and the proposed hybrid centrality measure (HADC, EADC( $\alpha = 0.25$ ), EADC( $\alpha = 0.5$ ), EADC( $\alpha = 0.75$ )) in the considered networks.

*Considered real-world data sets* the proposed solution to solve spread maximization problem discussed in this paper hybridizes degree centrality within closeness/harmonic and decay centrality. The considered problem is related to spread of trends, behaviors, influences, etc., therefore, we have picked 15 moderate size real-world social networks for simulations. We provide a brief summary of these networks in Table 7 and [31, 36] can be referred for a detailed description of the networks. The columns of Table 7 consist of names of the network instances, the number of nodes ( $n$ ), the number of edges ( $m$ ), the average degree of the networks (Avg. deg.), density of the network, average clustering coefficient ( $\hat{C}$ ), degree assortativity and the size of the maximum clique, respectively.

*Simulation* we have conducted simulations for evaluating the performance of various traditional centrality measures and the hybrid measures proposed in this paper for spread maximization problem which are summarized in Tables 8, 9, 10 and 11. The simulation is carried out as follows. The top-ranked node for every measure is taken as the source of the infection. The source and its neighborhood are infected at time  $t = 0$ . At each time step the infection is propagated as per the diffusion model being used. The contagion ends if it is unable to infect any susceptible nodes in a time-step. The total number of infected nodes by the end of the contagion is the spread-score for the source node in the simulation. The performance of a centrality measure is evaluated by computing the spread-score of the top most central node as per the centrality measure.

For the deterministic threshold model, the spread-scores of top most central nodes according to the traditional and proposed hybrid centrality measures in all considered real-world networks for different possible threshold value in the interval (0,1) with a step increase of 0.1 are plotted in Fig. 2. The figure shows the number of infections on the  $y$ -axis and the threshold values on the  $x$ -axis. As a specific example consider the plot for Twitter Lists. The number of infections have a large dip when the threshold increases from 0.1 to 0.2. The worst performing metric at the threshold of 0.2 is EC. When increasing the threshold from 0.2 to 0.3, we see another dip in number of infections, with DC and BC performing almost equally while BC was superior at the 0.2 threshold. CC and proposed centrality measures outperform the other measures across the rest of the thresholds. The main aim to plot these is to understand, for which interval of threshold values, the spread-scores for various centrality measures vary majorly. It is clear from the plots that for most of the considered real-world networks, the spread scores vary mostly in the sub-interval [0.3, 0.6]. Yet, the spread-scores for different measures are not clear and comparable from these plots. Therefore, the results of the simulation for threshold value 0.3, 0.4, 0.5, and 0.6, i.e., for threshold in terms of percentage set at 30%, 40%, 50% and 60% are presented in Tables 8, 9, 10, and 11, respectively. For each of the tables, the first column shows the network name, the second column shows the spread-scores of Betweenness Centrality, the third column shows the spread-scores of



Closeness Centrality, the fourth column shows the spread-scores for the Degree Centrality and the fifth column shows the spread-scores for Eigenvector Centrality. The other half of the table, from the sixth column onward shows the scores of the proposed hybrid measures. Namely, the sixth column shows the spread-score for HADC, the seventh

**Table 1** Considered real-world networks

Instance name	$n$	$m$	Avg. deg.	Density	$\hat{C}$	Assortativity (D)	Max clique	Network type
Chicago Road [36]	1467	1298	1.7696	0.0012	0.0001	-0.5049	2	Transport network
Chilean Power Grid [38]	347	444	2.5591	0.0074	0.0865	-0.0773	4	Energy network
Euroroad [36]	1174	1417	2.414	0.0021	0.0167	0.1267	3	Transport network
London Metro [34]	266	308	2.3158	0.0087	0.0363	0.1531	3	Transport network
Madrid Metro [34]	209	240	2.2967	0.011	0.0056	0.1776	3	Transport network
Mexico Metro [34]	147	164	2.2313	0.0153	0.0034	-0.1105	3	Transport network
Minnesota Road [31]	2642	3303	2.5004	0.0009	0.016	-0.1848	3	Transport network
Moscow Metro [34]	134	156	2.3284	0.0175	0.0174	0.4492	3	Transport network
New York Metro [34]	433	475	2.194	0.0051	0.0173	-0.0297	3	Transport network
Oldenburg Road [32]	6105	7029	2.3027	0.0004	0.0108	0.0554	3	Transport network
Openflights [34]	2939	15677	10.6683	0.0036	0.4526	0.0509	22	Transport network
Osaka Metro [34]	108	123	2.2778	0.0213	0.0001	0.1965	2	Transport network
p2p-Gnutella08 [37]	6301	20777	6.5948	0.001	0.0109	0.0356	5	Internet peer-to-peer network
Paris Metro [34]	299	356	2.3813	0.008	0.0204	-0.0297	3	Transport network
as20000102 [37]	6474	13895	4.2926	0.0007	0.2522	-0.1704	10	Autonomous systems graph
Seoul Metro [34]	392	437	2.2296	0.0057	0.006	0.0313	3	Transport network
Shanghai Metro [34]	148	158	2.1351	0.0145	0.0029	-0.0318	3	Transport network
Tokyo Metro [34]	217	262	2.4147	0.0112	0.0237	0.1983	3	Transport network
US Air [35]	332	2126	12.8072	0.0387	0.6252	-0.2079	22	Transport network
US Power Grid [39]	4941	6594	2.6691	0.0005	0.0801	0.0035	6	Energy network
NRPG data [33]	246	373	3.0325	0.0124	0.1071	-0.0932	3	Energy network

column for EADC with  $\alpha = 0.25$ , the eight column for EADC with  $\alpha = 0.5$  and the ninth column for EADC with  $\alpha = 0.75$ .

We have emphasized in bold, the best spread-score in Table 8, 9, 10, 11, 12 and 13 for each network instances given in Table 7 over various models, when the seed node is picked according to the considered centrality measures. As seen in Table 8, at a lower threshold, CC performs better than other traditional measures for most networks. However, as we increase the threshold, we observe that DC starts performing better than other traditional measures, as shown in Table 9. Since our measures are a hybrid of both the measures, they are able to maintain high scores across the thresholds. As an example if we look at Table 9 for Twitter Lists, the measures have similar results as CC, on the

**Table 2** Expected response-time of the proposed hybrid measures and the traditional centrality measures when the most central node is made the service center

Instance name	BC	DC	CC	EC	HABC	EABC ( $\alpha = 0.5$ )	EABC ( $\alpha = 0.75$ )
Chicago Road	<b>0.49</b>	<b>0.49</b>	<b>0.49</b>	<b>0.49</b>	<b>0.49</b>	<b>0.49</b>	<b>0.49</b>
Chilean Power Grid	<b>0.29</b>	0.33	<b>0.29</b>	0.33	<b>0.29</b>	<b>0.29</b>	<b>0.29</b>
Euroroad	<b>0.88</b>	<b>0.88</b>	0.89	1.05	0.88	<b>0.88</b>	<b>0.88</b>
London Metro	0.83	0.83	<b>0.80</b>	0.83	<b>0.80</b>	<b>0.80</b>	<b>0.80</b>
Madrid Metro	0.46	0.46	0.44	0.66	0.46	0.46	<b>0.42</b>
Mexico Metro	<b>0.50</b>	<b>0.50</b>	0.54	<b>0.50</b>	<b>0.50</b>	<b>0.50</b>	0.54
Minnesota Road	2.29	3.21	2.15	3.11	3.29	<b>2.29</b>	3.45
Moscow Metro	0.40	0.40	<b>0.37</b>	0.39	0.39	0.39	<b>0.37</b>
New York Metro	<b>1.11</b>	1.3	1.12	1.3	<b>1.11</b>	<b>1.11</b>	<b>1.11</b>
Oldenburg Road	<b>1.66</b>	<b>1.66</b>	1.68	2.64	<b>1.66</b>	<b>1.66</b>	<b>1.66</b>
Openflights	<b>0.16</b>	<b>0.16</b>	<b>0.16</b>	<b>0.16</b>	<b>0.16</b>	<b>0.16</b>	<b>0.16</b>
Osaka Metro	0.48	0.48	0.44	0.48	0.48	0.48	<b>0.42</b>
p2p-Gnutella08	<b>0.29</b>	<b>0.27</b>	<b>0.27</b>	<b>0.27</b>	<b>0.27</b>	<b>0.27</b>	<b>0.27</b>
Paris Metro	0.55	0.55	0.55	0.56	<b>0.51</b>	<b>0.51</b>	<b>0.51</b>
as20000102	<b>0.12</b>	<b>0.12</b>	<b>0.12</b>	<b>0.12</b>	<b>0.12</b>	<b>0.12</b>	<b>0.12</b>
Seoul Metro	1.06	1.25	1.06	1.23	<b>1.02</b>	<b>1.02</b>	<b>1.02</b>
Shanghai Metro	<b>0.53</b>	<b>0.53</b>	0.64	<b>0.53</b>	0.64	<b>0.53</b>	0.64
Tokyo Metro	0.57	0.57	0.57	0.57	0.57	0.57	<b>0.49</b>
US Air	<b>0.10</b>	<b>0.10</b>	<b>0.10</b>	<b>0.10</b>	<b>0.10</b>	<b>0.10</b>	<b>0.10</b>
US Power Grid	0.92	1.34	<b>0.83</b>	2.12	<b>0.83</b>	<b>0.83</b>	<b>0.83</b>
NRPG	<b>0.30</b>	0.45	<b>0.30</b>	0.37	<b>0.30</b>	<b>0.30</b>	<b>0.30</b>

other hand, in the same table for soc-pages-media and soc-gamesec-RO the results are similar to DC. It is evident from the tables that the proposed measures are a hybridization of DC and CC, and therefore, they perform at par with the better performing traditional measures across various thresholds. In some cases, the proposed measures even outperform all the traditional measures.

Apart from the deterministic linear threshold model, we have also evaluated the performance of the proposed measures as well as traditional measure using stochastic diffusion models, namely independent cascade (IC) and stochastic linear threshold (LT). The results of these experiments are shown in Tables 13 and 12, respectively. The results shown in the tables are the averages taken over 100 independent iterations of the diffusion models. From Table 12, we see that HADC and EADC with  $\alpha = 0.75$  have the highest number of expected infections across most networks. For some networks such as soc-pages-public-figure, they outperform all of the traditional measures. In the network Twitter lists, CC is the top performing traditional measure while in the network soc-pages-media, DC and EC are the top performing traditional measures. In both cases HADC and EADC with  $\alpha = 0.75$  perform optimally as they are a type of hybridization between CC and DC. In case of the IC model as shown in Table 13, we see that no measure performs consistently across the networks. As none of the traditional measures are able to perform consistently across the networks, a hybridization between them may lead to sub-optimal results.

We present the Spearman's rank correlation coefficient, Kendall's rank correlation coefficient and Fagin's intersection metric [40] to evaluate the correlation between traditional centrality measures and proposed hybrid centrality measures in Tables 14,

**Table 3 Ranking list of top three central nodes using the proposed hybrid measures and the traditional centrality measures on considered real-world networks given in Table 1**

Instance name	HABC			EABC ( $\alpha = 0.5$ )			EABC ( $\alpha = 0.75$ )			BC			CC			DC			EC		
	R1	R2	R3	R1	R2	R3	R1	R2	R3	R1	R2	R3	R1	R2	R3	R1	R2	R3	R1	R2	R3
Chicago Road	1156	1157	1147	1156	1157	1147	1156	1157	1147	1156	1157	1147	1156	1157	1147	552	922	10	1156	1147	1150
Chilean Power Grid	178	105	147	178	105	147	178	105	147	178	105	147	178	105	147	108	6	178	108	264	29
Euroroad	<b>402</b>	284	277	402	284	277	<b>402</b>	284	277	<b>402</b>	284	277	<b>402</b>	284	277	284	7	39	7	43	454
London Metro	<b>115</b>	29	12	115	29	12	12	115	180	<b>115</b>	29	214	12	164	305	214	115	219	214	115	219
Madrid Metro	<b>80</b>	162	146	80	162	146	<b>80</b>	146	162	<b>80</b>	146	162	162	80	235	80	15	9	15	54	40
Mexico Metro	<b>26</b>	70	23	26	70	78	<b>26</b>	70	54	<b>26</b>	70	25	70	26	23	26	127	61	26	70	23
Minnesota Road	300	328	280	1821	2069	300	1821	2069	300	1821	2069	2063	1356	1820	1109	2418	35	32	1927	1930	1987
Moscow Metro	59	71	25	59	71	25	73	25	74	59	73	71	25	24	59	21	59	21	59	25	73
New York Metro	<b>424</b>	191	410	424	191	410	<b>424</b>	191	410	<b>424</b>	191	148	84	424	410	273	228	410	273	228	188
Oldenburg Road	<b>831</b>	714	731	831	714	803	<b>831</b>	714	803	<b>831</b>	714	593	714	731	712	831	1657	1571	1677	1683	1691
Openflights	53	65	57	53	57	65	53	57	65	53	65	1576	53	65	57	53	65	59	53	59	65
Osaka Metro	<b>25</b>	70	95	25	70	95	<b>25</b>	99	70	<b>25</b>	99	70	95	70	25	25	54	42	25	81	54
p2p-Gnutella08	<b>123</b>	367	127	123	367	127	123	367	5831	<b>123</b>	367	424	<b>123</b>	127	1317	123	127	367	367	123	145
Paris Metro	150	57	235	150	57	235	150	57	235	150	57	106	57	150	235	57	177	246	186	150	235
as20000102	2	10	7	2	10	7	2	7	10	2	7	10	2	10	7	2	10	7	2	10	7
Seoul Metro	220	480	357	220	444	391	220	444	391	220	444	391	220	480	357	254	444	391	103	254	125
Shanghai Metro	<b>136</b>	16	113	136	16	213	16	136	65	<b>136</b>	113	97	16	136	220	16	136	220	16	137	138
Tokyo Metro	150	85	53	150	85	53	150	60	150	150	60	85	150	60	85	150	53	120	150	120	36
US Air	118	261	201	118	201	47	118	8	261	118	8	261	118	261	67	118	261	255	118	261	255
US Power Grid	<b>1378</b>	1365	2685	1378	1365	2685	1378	1365	1678	651	559	1365	<b>1378</b>	1678	2944	2847	602	932	4422	4436	4419
NRP data	<b>233</b>	235	238	233	235	238	233	235	238	70	235	213	<b>233</b>	238	70	194	65	56	239	213	215

**Table 4 Spearman's rank correlation between the proposed and the traditional centrality measures and on considered real-world networks given in Table 1**

Instance name	HABC				EABC ( $\alpha = 0.5$ )				EABC ( $\alpha = 0.75$ )			
	BC	CC	DC	EC	BC	CC	DC	EC	BC	CC	DC	EC
	Chicago Road	0.2087	0.3688	0.2393	0.2027	0.2040	0.3437	0.2351	0.2024	0.2035	0.3241	0.1985
Chilean Power Grid	0.0691	-0.0620	0.0132	0.0057	0.0224	0.0274	0.0506	0.0163	0.0929	0.0867	-0.0200	0.0850
Euroroad	0.7707	0.0144	0.2019	-0.0010	0.7985	0.0190	0.1675	-0.0174	0.7656	0.0214	0.1834	-0.0227
London Metro	0.1625	0.1450	0.0296	0.0312	-0.0291	0.0323	0.0917	0.0636	0.0432	0.1496	-0.0837	0.0324
Madrid Metro	0.0540	0.1561	-0.0912	0.0149	0.0284	0.0891	-0.0949	0.1213	0.0109	0.0814	-0.1316	0.0049
Mexico Metro	-0.0598	0.2043	0.0428	0.1145	0.1267	-0.0479	0.0516	0.1401	-0.0037	0.2043	-0.0716	0.1522
Minnesota Road	-0.0068	-0.0358	-0.0777	-0.1648	0.0286	-0.0221	0.0075	-0.0393	0.0015	-0.0552	0.0345	-0.1444
Moscow Metro	0.0824	0.0474	0.1245	0.0570	-0.1814	-0.1215	-0.2326	0.0627	0.1012	0.0468	0.0331	-0.1045
New York Metro	0.0386	0.0865	-0.0061	0.0760	0.0599	0.0394	-0.0382	-0.0222	0.0083	-0.0139	-0.0651	-0.0156
Oldenburg Road	0.0098	0.1332	0.0094	-0.0222	0.0075	0.0593	0.0202	0.0169	0.0113	0.0991	0.0039	-0.0083
Openflights	0.0018	0.1090	0.0263	0.0799	0.0055	0.1417	0.0318	0.0898	0.0032	0.1305	0.0200	0.0618
Osaka Metro	0.0110	0.1728	0.0262	0.1440	0.0482	0.1662	0.0402	0.1398	0.0795	0.1031	-0.0713	0.0656
p2p-Gnutella08	0.0537	0.0760	0.0880	0.0491	0.0636	0.0634	0.0775	0.0485	0.0660	0.0529	0.0754	0.0561
as20000102	0.1327	0.5878	0.1837	0.4513	0.1408	0.5881	0.1810	0.4423	0.1309	0.6273	0.1912	0.4392
Seoul Metro	0.0520	-0.0055	-0.0048	-0.0389	0.0257	0.0143	0.0480	0.0775	-0.0427	0.0109	-0.0643	0.0136
Shanghai Metro	0.1104	-0.0165	0.1168	0.0977	0.0721	0.1124	0.0326	0.1617	0.0014	0.0922	0.1809	0.0610
US Air	0.0855	0.1707	0.0522	0.0608	0.1171	0.2246	0.0059	0.0674	0.0503	0.2372	0.0243	0.1065
US Power Grid	0.0450	0.0831	0.0830	0.0437	0.0350	0.0279	0.0345	0.0038	0.0436	0.0598	0.0622	0.0436
Paris Metro	-0.1077	0.0421	-0.0625	0.1169	0.0302	0.0846	-0.0271	0.0351	0.0175	0.1305	-0.0914	0.0309
Tokyo Metro	0.1693	0.0920	0.0064	0.0460	0.0555	0.0883	0.0958	0.0511	0.0646	0.1071	-0.0794	0.1039
NRPG data	0.0118	0.1334	-0.0693	0.0934	0.0445	0.0975	-0.1389	0.1395	-0.0867	0.1659	-0.1322	0.0936
Average correlation	0.0902	0.1192	0.0444	0.0694	0.0811	0.0966	0.0305	0.0858	0.0744	0.1268	0.0094	0.0574
Standard deviation	0.1729	0.1441	0.0907	0.1152	0.1802	0.1485	0.1043	0.1042	0.1701	0.1432	0.1069	0.1139

**Table 5 Kendall rank correlation coefficient (Kendall's Tau) between the proposed and the traditional centrality measures and on considered real-world networks given in Table 1**

Instance name	HABC				EABC ( $\alpha = 0.5$ )				EABC ( $\alpha = 0.75$ )			
	BC	CC	DC	EC	BC	CC	DC	EC	BC	CC	DC	EC
Chicago Road	0.2402	0.2919	0.2531	0.1375	0.2493	0.2762	0.2601	0.1197	0.2375	0.2669	0.2267	0.1024
Chilean Power Grid	0.0461	-0.0379	0.0089	0.0045	0.0149	0.0197	0.0318	0.0119	0.0621	0.0651	-0.0153	0.0556
Euroroad	0.7126	0.0099	0.1404	-0.0004	0.7461	0.0127	0.1168	-0.0115	0.7025	0.0147	0.1278	-0.0149
London Metro	0.1074	0.0972	0.0173	0.0220	-0.0200	0.0241	0.0617	0.0445	0.0286	0.1022	-0.0568	0.0241
Madrid Metro	0.0334	0.1058	-0.0609	0.0098	0.0215	0.0595	-0.0647	0.0846	0.0097	0.0558	-0.0885	0.0077
Mexico Metro	-0.0445	0.1386	0.0264	0.0724	0.0860	-0.0351	0.0338	0.1026	-0.0040	0.1376	-0.0566	0.1047
Minnesota Road	-0.0047	-0.0216	-0.0432	-0.1205	0.0190	-0.0143	0.0052	-0.0264	0.0012	-0.0370	0.0232	-0.0999
Moscow Metro	0.0553	0.0315	0.0809	0.0376	-0.1274	-0.0807	-0.1592	0.0439	0.0690	0.0349	0.0250	-0.0722
New York Metro	0.0259	0.0582	-0.0040	0.0501	0.0399	0.0265	-0.0254	-0.0143	0.0067	-0.0095	-0.0429	-0.0094
Oldenburg Road	0.0066	0.0905	0.0067	-0.0146	0.0051	0.0401	0.0136	0.0118	0.0074	0.0671	0.0029	-0.0056
Openflights	0.0053	0.0797	0.0217	0.0571	0.0077	0.1029	0.0249	0.0645	0.0061	0.0936	0.0184	0.0443
Osaka Metro	0.0097	0.1222	0.0194	0.1038	0.0301	0.1177	0.0280	0.1000	0.0568	0.0723	-0.0519	0.0561
p2p-Gnutella08	0.0364	0.0514	0.0595	0.0329	0.0433	0.0427	0.0524	0.0326	0.0447	0.0360	0.0513	0.0379
as20000102	0.0909	0.4913	0.1318	0.3756	0.0963	0.4917	0.1303	0.3696	0.0900	0.5191	0.1367	0.3672
Seoul Metro	0.0349	-0.0052	-0.0031	-0.0280	0.0165	0.0090	0.0305	0.0527	-0.0281	0.0056	-0.0439	0.0112
Shanghai Metro	0.0761	-0.0097	0.0798	0.0689	0.0506	0.0735	0.0182	0.1158	0.0029	0.0671	0.1129	0.0421
US Air	0.0568	0.1184	0.0345	0.0401	0.0782	0.1584	0.0051	0.0473	0.0339	0.1685	0.0186	0.0717
US Power Grid	0.0304	0.0559	0.0558	0.0294	0.0234	0.0186	0.0229	0.0025	0.0292	0.0406	0.0418	0.0289
Paris Metro	-0.0721	0.0269	-0.0417	0.0795	0.0186	0.0589	-0.0180	0.0217	0.0119	0.0928	-0.0620	0.0184
Tokyo Metro	0.1150	0.0610	0.0063	0.0311	0.0384	0.0608	0.0626	0.0315	0.0445	0.0742	-0.0559	0.0707
NRPG data	0.0069	0.0888	-0.0416	0.0598	0.0283	0.0660	-0.0889	0.0972	-0.0596	0.1109	-0.0914	0.0625
Average correlation	0.0747	0.0878	0.0356	0.0499	0.0698	0.0728	0.0258	0.0620	0.0644	0.0942	0.0105	0.0430
Standard deviation	0.1594	0.1167	0.0729	0.0911	0.1684	0.1205	0.0833	0.0830	0.1572	0.1172	0.0835	0.0895

**Table 6** Fagin's intersection metric between the proposed and the traditional centrality measures and on considered real-world networks given in Table 1

Instance name	HABC				EABC ( $\alpha = 0.5$ )				EABC ( $\alpha = 0.75$ )			
	BC	CC	DC	EC	BC	CC	DC	EC	BC	CC	DC	EC
	Chicago Road	0.6487	0.8122	0.6106	0.8187	0.6639	0.7809	0.6286	0.7873	0.6308	0.8073	0.5992
Chilean Power Grid	0.7007	0.9529	0.6461	0.7418	0.7105	0.9212	0.6553	0.7361	0.6912	0.966	0.6398	0.7416
Euroroad	0.9986	0.5866	0.6477	0.4884	0.9992	0.5869	0.648	0.4888	0.9983	0.5863	0.6477	0.4882
London Metro	0.6801	0.9379	0.6831	0.8843	0.7217	0.8795	0.7013	0.8351	0.6582	0.9546	0.6729	0.8954
Madrid Metro	0.7071	0.9205	0.6645	0.7385	0.749	0.8651	0.6862	0.6978	0.6809	0.9407	0.6542	0.7437
Mexico Metro	0.7546	0.9159	0.6332	0.9014	0.7898	0.8676	0.6409	0.858	0.7289	0.9326	0.618	0.9006
Minnesota Road	0.2845	0.1173	0.3375	0.0239	0.6407	0.2839	0.3661	0.1631	0.3023	0.1223	0.3434	0.0615
Moscow Metro	0.7809	0.9413	0.7758	0.9056	0.7967	0.9193	0.7825	0.8875	0.7615	0.9596	0.7769	0.9083
New York Metro	0.704	0.9081	0.6248	0.7317	0.7949	0.7963	0.6488	0.658	0.7004	0.9022	0.6253	0.7257
Oldenburg Road	0.493	0.8842	0.1815	0.1423	0.7121	0.6117	0.2158	0.1234	0.5207	0.8129	0.1877	0.129
Openflights	0.5825	0.9675	0.7762	0.8194	0.5798	0.9674	0.7753	0.8207	0.5765	0.9632	0.7704	0.8194
Osaka Metro	0.7505	0.9264	0.6979	0.869	0.7758	0.8949	0.7085	0.8427	0.7134	0.958	0.6749	0.8659
p2p-Gnutella08	0.4044	0.926	0.5036	0.8427	0.3999	0.9203	0.4997	0.8474	0.4049	0.9291	0.5057	0.8394
as20000102	0.3204	0.9779	0.4107	0.8947	0.3181	0.9792	0.4085	0.8932	0.3245	0.9823	0.4143	0.8864
Seoul Metro	0.6927	0.9099	0.6274	0.709	0.7945	0.794	0.6263	0.641	0.6883	0.9047	0.6269	0.7098
Shanghai Metro	0.7429	0.8944	0.6196	0.7672	0.7999	0.8264	0.6534	0.7504	0.7165	0.92	0.6016	0.7629
US Air	0.8097	0.9562	0.8614	0.8836	0.8056	0.9541	0.8574	0.8823	0.8026	0.951	0.8538	0.8786
US Power Grid	0.5087	0.8876	0.2427	0.1926	0.5869	0.7553	0.2679	0.1789	0.5025	0.8659	0.2449	0.1949
Paris Metro	0.7242	0.9392	0.6342	0.9106	0.7554	0.8876	0.6523	0.8702	0.7059	0.955	0.6258	0.9207
Tokyo Metro	0.7012	0.932	0.6779	0.8424	0.7305	0.8956	0.69	0.8181	0.6758	0.9439	0.6631	0.843
NRFG data	0.8047	0.9426	0.6834	0.7305	0.8181	0.917	0.694	0.7396	0.7833	0.9572	0.6749	0.7213
Average correlation	0.6569	0.8684	0.5971	0.7066	0.7116	0.8240	0.6098	0.6914	0.6461	0.8721	0.5915	0.7073
Standard deviation	0.1715	0.1903	0.1716	0.2655	0.1480	0.1613	0.1659	0.2452	0.1629	0.1929	0.1671	0.2621

**Table 7** Considered real-world social networks

Instance name	$n$	$m$	Avg. deg.	Density	$\hat{c}$	Assortativity (D)	Max clique
soc-gplus [31]	23,628	39,242	3.3217	0.0001	0.0037	-0.387	7
Twitter Lists [36]	23,370	33,101	2.8328	0.0001	0.0215	-0.4741	7
soc-hamsterer [31]	2426	16,630	13.7098	0.0057	0.2314	0.0474	25
soc-pages-government [31]	7057	89,455	25.3521	0.0036	0.2238	0.0294	28
soc-pages-politician [31]	5908	41,729	14.1263	0.0024	0.3011	0.0184	21
soc-pages-shows [31]	3892	17,262	8.8705	0.0023	0.5906	0.5605	57
soc-wiki-elec [31]	7118	107,071	30.0846	0.0042	0.1255	-0.0514	17
soc-advogato [31]	6551	51,332	15.6715	0.0024	0.0925	-0.0526	19
soc-pages-public-figure [31]	11,565	67,114	11.6064	0.001	0.1666	0.2024	21
soc-anybeat [31]	12,645	67,053	10.6055	0.0008	0.0217	-0.1219	25
soc-pages-sport [31]	13,866	86,858	12.5282	0.0009	0.1292	-0.0268	29
soc-epinions [31]	26,588	100,120	7.5312	0.0003	0.0892	0.0572	16
soc-pages-media [31]	27,917	206,259	14.7766	0.0005	0.114	0.0221	31
soc-pages-company [31]	14,113	52,310	7.413	0.0005	0.1532	0.0135	21
soc-gamesec-RO [31]	41,773	125,826	6.0243	0.0001	0.0753	0.114	7

**Table 8** Number of infected nodes when the top-ranked node is chosen as the source of the infection for threshold = 30%

Instance name	BC	CC	DC	EC	HADC	EADC ( $\alpha = 0.25$ )	EADC ( $\alpha = 0.5$ )	EADC ( $\alpha = 0.75$ )
soc-gplus	<b>1,4951</b>	<b>14,951</b>	<b>14,951</b>	<b>14,951</b>	<b>14,951</b>	<b>14,951</b>	<b>14,951</b>	<b>14,951</b>
Twitter Lists	484	<b>1180</b>	553	539	<b>1180</b>	<b>1180</b>	<b>1180</b>	<b>1180</b>
soc-hamsterer	<b>1879</b>	<b>1879</b>	<b>1879</b>	<b>1879</b>	<b>1879</b>	<b>1879</b>	<b>1879</b>	<b>1879</b>
soc-pages-government	<b>4316</b>	<b>4316</b>	<b>4316</b>	3695	<b>4316</b>	<b>4316</b>	<b>4316</b>	<b>4316</b>
soc-pages-politician	<b>2185</b>	<b>2185</b>	946	674	<b>2185</b>	<b>2185</b>	<b>2185</b>	<b>2185</b>
soc-pages-shows	<b>1283</b>	<b>1283</b>	1242	210	<b>1283</b>	279	<b>1283</b>	<b>1283</b>
soc-wiki-elec	<b>7061</b>	<b>7061</b>	<b>7061</b>	<b>7061</b>	<b>7061</b>	<b>7061</b>	<b>7061</b>	<b>7061</b>
soc-advogato	<b>5014</b>	<b>5014</b>	<b>5014</b>	<b>5014</b>	<b>5014</b>	<b>5014</b>	<b>5014</b>	<b>5014</b>
soc-pages	10,095	10,095	10,029	1485	<b>10,184</b>	1485	<b>10,184</b>	<b>10,184</b>
soc-anybeat	<b>12,538</b>	<b>12,538</b>	<b>12,538</b>	<b>12,538</b>	<b>12,538</b>	<b>12,538</b>	<b>12,538</b>	<b>12,538</b>
soc-pages-sport	<b>13,739</b>	<b>13,739</b>	<b>13,739</b>	<b>13,739</b>	<b>13,739</b>	<b>13,739</b>	<b>13,739</b>	<b>13,739</b>
soc-epinions	2094	3106	<b>25,283</b>	<b>25,283</b>	<b>25,283</b>	<b>25,283</b>	<b>25,283</b>	3106
soc-pages-media	<b>26,982</b>	26,981	26,981	26,981	26,981	26,981	26,981	26,981
soc-pages-company	1065	<b>12,432</b>	<b>12,432</b>	163	<b>12,432</b>	<b>12,432</b>	<b>12,432</b>	<b>12,432</b>
soc-gamesec-RO	219	219	554	<b>555</b>	219	554	219	219

15 and 16, respectively. For our purposes we compute the Fagin's rank intersection for top-1000 ranks in the network. In case the number of nodes is less than or equal to 1000, the rank intersection is computed for all nodes in the network. For each of tables, the first column shows the Network Name. The next four columns show the

**Table 9** Number of infected nodes when the top-ranked node is chosen as the source of the infection for threshold = 40%

Instance name	BC	CC	DC	EC	HADC	EADC ( $\alpha = 0.25$ )	EADC ( $\alpha = 0.5$ )	EADC ( $\alpha = 0.75$ )
soc-gplus	<b>14,556</b>	<b>14,556</b>	<b>14,556</b>	<b>14,556</b>	<b>14,556</b>	<b>14,556</b>	<b>14,556</b>	<b>14,556</b>
Twitter Lists	473	<b>912</b>	452	522	<b>912</b>	<b>912</b>	<b>912</b>	<b>912</b>
soc-hamsterer	<b>1666</b>	<b>1666</b>	<b>1666</b>	<b>1666</b>	<b>1666</b>	<b>1666</b>	<b>1666</b>	<b>1666</b>
soc-pages-government	<b>3372</b>	<b>3372</b>	<b>3372</b>	1446	<b>3372</b>	<b>3372</b>	<b>3372</b>	<b>3372</b>
soc-pages-politician	552	552	<b>652</b>	387	552	552	552	552
soc-pages-shows	297	297	<b>299</b>	141	297	196	297	297
soc-wiki-elec	<b>7061</b>	<b>7061</b>	<b>7061</b>	<b>7061</b>	<b>7061</b>	<b>7061</b>	<b>7061</b>	<b>7061</b>
soc-advogato	<b>4667</b>	<b>4667</b>	<b>4667</b>	<b>4667</b>	<b>4667</b>	<b>4667</b>	<b>4667</b>	<b>4667</b>
soc-pages	1143	1143	<b>7825</b>	1328	1223	1328	1223	1223
soc-anybeat	<b>12,287</b>	<b>12,287</b>	<b>12,287</b>	<b>12,287</b>	<b>12,287</b>	<b>12,287</b>	<b>12287</b>	<b>12287</b>
soc-pages-sport	<b>1082</b>	<b>1082</b>	<b>1082</b>	<b>1082</b>	<b>1082</b>	<b>1082</b>	<b>1082</b>	<b>1082</b>
soc-epinions	1383	1824	<b>22,581</b>	<b>22,581</b>	<b>22,581</b>	<b>22,581</b>	<b>22,581</b>	1824
soc-pages-media	1101	1189	<b>1694</b>	<b>1694</b>	<b>1694</b>	<b>1694</b>	<b>1694</b>	<b>1694</b>
soc-pages-company	467	<b>561</b>	<b>561</b>	121	<b>561</b>	<b>561</b>	<b>561</b>	<b>561</b>
soc-gamesec-RO	129	129	<b>180</b>	175	129	<b>180</b>	129	129

**Table 10** Number of infected nodes when the top-ranked node is chosen as the source of the infection for threshold = 50%

Instance name	BC	CC	DC	EC	HADC	EADC ( $\alpha = 0.25$ )	EADC ( $\alpha = 0.5$ )	EADC ( $\alpha = 0.75$ )
soc-gplus	<b>11,482</b>	<b>11,482</b>	<b>11,482</b>	<b>11,482</b>	<b>11,482</b>	<b>11,482</b>	<b>11,482</b>	<b>11,482</b>
Twitter Lists	473	<b>850</b>	452	522	<b>850</b>	<b>850</b>	<b>850</b>	<b>850</b>
soc-hamsterer	<b>1091</b>	<b>1091</b>	<b>1091</b>	<b>1091</b>	<b>1091</b>	<b>1091</b>	<b>1091</b>	<b>1091</b>
soc-pages-government	<b>2631</b>	<b>2631</b>	<b>2631</b>	1094	<b>2631</b>	<b>2631</b>	<b>2631</b>	<b>2631</b>
soc-pages-politician	483	483	<b>640</b>	380	483	483	483	483
soc-pages-shows	<b>218</b>	<b>218</b>	216	137	<b>218</b>	172	<b>218</b>	<b>218</b>
soc-wiki-elec	<b>7061</b>	<b>7061</b>	<b>7061</b>	<b>7061</b>	<b>7061</b>	<b>7061</b>	<b>7061</b>	<b>7061</b>
soc-advogato	<b>4584</b>	<b>4584</b>	<b>4584</b>	<b>4584</b>	<b>4584</b>	<b>4584</b>	<b>4584</b>	<b>4584</b>
soc-pages-public-figure	905	905	958	<b>1242</b>	831	<b>1242</b>	831	831
soc-anybeat	<b>12,221</b>	<b>12,221</b>	<b>12,221</b>	<b>12,221</b>	<b>12,221</b>	<b>12,221</b>	<b>12,221</b>	<b>12,221</b>
soc-pages-sport	<b>837</b>	<b>837</b>	<b>837</b>	<b>837</b>	<b>837</b>	<b>837</b>	<b>837</b>	<b>837</b>
soc-epinions	1310	<b>1715</b>	1709	1709	1709	1709	1709	<b>1715</b>
soc-pages-media	911	984	<b>1159</b>	<b>1159</b>	<b>1159</b>	<b>1159</b>	<b>1159</b>	<b>1159</b>
soc-pages-company	417	<b>492</b>	<b>492</b>	101	<b>492</b>	<b>492</b>	<b>492</b>	<b>492</b>
soc-gamesec-RO	124	124	158	<b>159</b>	124	158	124	124

**Table 11** Number of infected nodes when the top-ranked node is chosen as the source of the infection for threshold = 60%

Instance name	BC	CC	DC	EC	HADC	EADC ( $\alpha = 0.25$ )	EADC ( $\alpha = 0.5$ )	EADC ( $\alpha = 0.75$ )
soc-gplus	<b>8741</b>	<b>8741</b>	<b>8741</b>	<b>8741</b>	<b>8741</b>	<b>8741</b>	<b>8741</b>	<b>8741</b>
Twitter Lists	402	<b>633</b>	417	500	<b>633</b>	<b>633</b>	<b>633</b>	<b>633</b>
soc-hamsterer	<b>567</b>	<b>567</b>	<b>567</b>	<b>567</b>	<b>567</b>	<b>567</b>	<b>567</b>	<b>567</b>
soc-pages-government	<b>1712</b>	<b>1712</b>	<b>1712</b>	944	<b>1712</b>	<b>1712</b>	<b>1712</b>	<b>1712</b>
soc-pages-politician	334	334	<b>558</b>	340	334	334	334	334
soc-pages-shows	<b>161</b>	<b>161</b>	<b>161</b>	124	<b>161</b>	150	<b>161</b>	<b>161</b>
soc-wiki-elec	<b>5241</b>	<b>5241</b>	<b>5241</b>	<b>5241</b>	<b>5241</b>	<b>5241</b>	<b>5241</b>	<b>5241</b>
soc-advogato	<b>1795</b>	<b>1795</b>	<b>1795</b>	<b>1795</b>	<b>1795</b>	<b>1795</b>	<b>1795</b>	<b>1795</b>
soc-pages-public-figure	498	498	570	<b>1062</b>	508	<b>1062</b>	508	508
soc-anybeat	<b>10130</b>	<b>10130</b>	<b>10130</b>	<b>10130</b>	<b>10130</b>	<b>10130</b>	<b>10130</b>	<b>10130</b>
soc-pages-sport	<b>633</b>	<b>633</b>	<b>633</b>	<b>633</b>	<b>633</b>	<b>633</b>	<b>633</b>	<b>633</b>
soc-epinions	897	<b>1107</b>	981	981	981	981	981	<b>1107</b>
soc-pages-media	640	724	<b>863</b>	<b>863</b>	<b>863</b>	<b>863</b>	<b>863</b>	<b>863</b>
soc-pages-company	275	<b>326</b>	<b>326</b>	95	<b>326</b>	<b>326</b>	<b>326</b>	<b>326</b>
soc-gamesec-RO	101	101	119	<b>121</b>	101	119	101	101

respective correlation between HADC and the traditional measures, namely BC, CC, DC and EC. The next four columns, from columns 6 to 9 show the respective correlation between EADC with  $\alpha = 0.25$  and the traditional measures, BC, CC, DC and EC. Finally the columns from 10 to 13 show the the respective correlation between EADC with  $\alpha = 0.75$  and BC, CC, DC and EC.

From the Spearman's and Kendall's rank correlation coefficients, we observe that our measures have a small positive correlation with traditional measures. The correlation is small enough to conclude that the proposed measures rank nodes differently from traditional measures. The correlations are positive and almost same for all traditional measures. This indicates that our measures share some characteristics with them and are able to switch between traditional measures to perform optimally across networks and thresholds. Fagin's intersection metric has a higher value due to only considering top-1000 ranked nodes to compute the metric.

**Table 12** Expected number of infected nodes when the top-ranked node is chosen as the source of the infection for stochastic linear threshold model

Instance name	BC	CC	DC	EC	HADC	EADC ( $\alpha = 0.25$ )	EADC ( $\alpha = 0.5$ )	EADC ( $\alpha = 0.75$ )
soc-gplus	<b>11,990.96</b>	<b>11,990.96</b>	<b>11,990.96</b>	<b>11,990.96</b>	<b>11,990.96</b>	<b>11,990.96</b>	<b>11,990.96</b>	<b>11,990.96</b>
Twitter Lists	671.22	<b>1298.42</b>	697.3	592.26	<b>1298.42</b>	<b>1298.42</b>	<b>1298.42</b>	<b>1298.42</b>
soc-hamsterer	<b>1242.72</b>	<b>1242.72</b>	<b>1242.72</b>	<b>1242.72</b>	<b>1242.72</b>	<b>1242.72</b>	<b>1242.72</b>	<b>1242.72</b>
soc-pages-government	<b>4027.29</b>	<b>4027.29</b>	<b>4027.29</b>	<b>2916.12</b>	<b>4027.29</b>	<b>4027.29</b>	<b>4027.29</b>	<b>4027.29</b>
soc-pages-politician	<b>1907.98</b>	<b>1907.98</b>	1123.45	724.17	<b>1907.98</b>	<b>1907.98</b>	<b>1907.98</b>	<b>1907.98</b>
soc-pages-shows	<b>748.47</b>	<b>748.47</b>	743.79	253.3	<b>748.47</b>	347.11	<b>748.47</b>	<b>748.47</b>
soc-wiki-elec	<b>6623.02</b>	<b>6623.02</b>	<b>6623.02</b>	<b>6623.02</b>	<b>6623.02</b>	<b>6623.02</b>	<b>6623.02</b>	<b>6623.02</b>
soc-advogato	<b>3741.1</b>	<b>3741.1</b>	<b>3741.1</b>	<b>3741.1</b>	<b>3741.1</b>	<b>3741.1</b>	<b>3741.1</b>	<b>3741.1</b>
soc-pages-public-figure	4415.09	4415.09	3928.15	2025.33	<b>4535.9</b>	2025.33	<b>4535.9</b>	<b>4535.9</b>
soc-anybeat	<b>11,394.31</b>	<b>11,394.31</b>	<b>11,394.31</b>	<b>11,394.31</b>	<b>11,394.31</b>	<b>11,394.31</b>	<b>11,394.31</b>	<b>11,394.31</b>
soc-pages-sport	<b>4064.69</b>	<b>4064.69</b>	<b>4064.69</b>	<b>4064.69</b>	<b>4064.69</b>	<b>4064.69</b>	<b>4064.69</b>	<b>4064.69</b>
soc-epinions	6660.87	<b>8353.97</b>	7469.31	7469.31	7469.31	7469.31	7469.31	<b>8353.97</b>
soc-pages-media	7927.39	10,118.52	<b>10,916.34</b>	<b>10,916.34</b>	<b>10,916.34</b>	<b>10,916.34</b>	<b>10,916.34</b>	<b>10,916.34</b>
soc-pages-company	1644.77	<b>2429.62</b>	<b>2429.62</b>	161.55	<b>2429.62</b>	<b>2429.62</b>	<b>2429.62</b>	<b>2429.62</b>
soc-gamesec-RO	1101.8	1101.8	<b>1293.95</b>	1235.01	1101.8	<b>1293.95</b>	1101.8	1101.8

**Table 13** Expected number of infected nodes when the top ranked node is chosen as the source of the infection for independent cascade model

Instance name	BC	CC	DC	EC	HADC	EADC ( $\alpha = 0.25$ )	EADC ( $\alpha = 0.5$ )	EADC ( $\alpha = 0.75$ )
soc-gplus	14,487.1	14,472.36	14,481.01	14,480.63	14,472.88	14,480.8	<b>14,487.43</b>	14,477.27
Twitter Lists	11,989.12	11,955.72	<b>11,992.85</b>	11,964.66	11,906.41	11,917.2	11,916.64	11,908.19
soc-hamsterer	937.39	937.74	<b>938.05</b>	936.18	937.11	937.69	936.21	937.34
soc-pages-government	3148.14	3149.21	3148.54	3077.98	3150.89	<b>3152.66</b>	3140.53	3144.39
soc-pages-politician	2771.62	2769.67	2640.67	2586.06	2770.33	2771.98	2769.76	<b>2774.51</b>
soc-pages-shows	1437.09	1442.95	1431.11	1308.85	1444.45	1330.98	1445.35	<b>1446.3</b>
soc-wiki-elec	2060.71	<b>2062.82</b>	2061.5	2061.38	2061.32	2061.22	2061.67	2061.75
soc-advogato	<b>2508.59</b>	2504.19	2505.09	2503.53	2504.48	2506.25	2505.2	2503.19
soc-pages-public-figure	4951.71	4949.31	<b>4955.46</b>	4792.01	4938.83	4792.36	4946.97	4940.21
soc-anybeat	5989.9	5992.17	5990.56	5989.8	5989.71	5991.55	<b>5994.93</b>	5991.79
soc-pages-sport	7147.25	7147.66	7142.17	7149.91	7146.72	<b>7150</b>	7149.07	7143.33
soc-epinions	<b>8131.74</b>	8083.85	7971.35	7966.32	7972.66	7963.08	7967.07	8074.81
soc-pages-media	<b>13272.46</b>	13217.94	13071.14	13069.32	13068.38	13071.39	13067.68	13075.13
soc-pages-company	<b>7085.8</b>	7027.44	7036.72	6951.13	7038.37	7033.11	7033.63	7030.34
soc-gamesec-RO	23,789.15	23,782.03	23,757.48	23,762.35	<b>23,795.5</b>	23,748.52	23,793.17	23,783.46

**Table 14 Spearman's rank correlation between the proposed and the traditional centrality measures and on considered real-world networks given in Table 7**

Instance name	HADC				EADC( $\alpha = 0.25$ )				EADC( $\alpha = 0.75$ )			
	BC	CC	DC	EC	BC	CC	DC	EC	BC	CC	DC	EC
	soc-gplus	0.2877	0.17	0.1634	0.0384	0.2871	0.2403	0.1921	0.173	0.2725	0.172	0.1522
Twitter Lists	0.0645	0.0315	0.0898	-0.0034	0.0639	0.0169	0.099	0.0837	0.0727	0.0661	0.102	-0.0006
soc-hamsterer	0.6595	0.7512	0.5876	0.6986	0.6472	0.7308	0.5757	0.6972	0.6636	0.7626	0.5742	0.6985
soc-pages-government	0.2753	0.296	0.3775	0.2483	0.2745	0.2877	0.3669	0.2698	0.2533	0.2903	0.3761	0.2698
soc-pages-politician	0.2079	0.197	0.2818	0.181	0.222	0.2039	0.3028	0.1982	0.182	0.1879	0.2883	0.1859
soc-pages-shows	0.1713	0.199	0.274	0.1608	0.2047	0.2167	0.3236	0.1685	0.1708	0.1917	0.2544	0.1736
soc-wiki-elec	-0.0191	0.0974	-0.0035	0.064	-0.0015	0.082	0.0027	0.0796	0.0003	0.0938	0.0087	0.0562
soc-advogato	0.4419	0.4772	0.399	0.47	0.4399	0.4628	0.3957	0.4805	0.4412	0.466	0.3936	0.4725
soc-pages	0.2768	0.2789	0.3643	0.2843	0.2943	0.2847	0.4008	0.2936	0.2763	0.2787	0.3667	0.2821
soc-anybeat	0.2006	0.249	0.1857	0.1463	0.1947	0.2019	0.1701	0.1928	0.2	0.2239	0.1968	0.1547
soc-pages-sport	0.2465	0.2275	0.2731	0.2186	0.2707	0.2239	0.2965	0.2188	0.2439	0.2202	0.2808	0.2081
soc-opinions	0.5778	0.7117	0.6392	0.7523	0.5796	0.7237	0.6466	0.7593	0.58	0.7151	0.6388	0.7506
soc-pages-media	0.2499	0.2552	0.3333	0.2567	0.2414	0.2564	0.3372	0.257	0.2411	0.2528	0.3287	0.2521
soc-pages-company	0.2085	0.2026	0.2759	0.1866	0.2353	0.2012	0.3035	0.1905	0.2208	0.184	0.2797	0.186
soc-wiki-Vote	0.0427	0.0754	0.033	0.1052	0.0553	0.0186	0.0237	0.0325	0.0444	0.0915	0.0764	0.0259
soc-gamesec-RO	0.1859	0.1727	0.2366	0.1531	0.1956	0.1738	0.2495	0.1644	0.1821	0.1596	0.2358	0.1644
Average correlation	0.25	0.27	0.28	0.25	0.26	0.27	0.29	0.27	0.25	0.27	0.28	0.25
Standard deviation	0.179	0.2052	0.175	0.2169	0.173	0.2078	0.1746	0.2074	0.1775	0.2051	0.1673	0.2183

**Table 15 Kendall rank correlation between the proposed and the traditional centrality measures and on considered real-world networks given in Table 7**

Instance name	HADC			EADC ( $\alpha = 0.25$ )			EADC ( $\alpha = 0.75$ )					
	BC	CC	EC	BC	CC	EC	BC	CC	EC			
	soc-gplus	0.1793	0.2636	0.0768	0.1842	0.2025	0.2167	0.1011	0.2355	0.1839	0.2248	0.0778
Twitter Lists	0.0607	0.0461	0.0450	0.0188	0.0563	0.0635	0.0497	-0.0099	0.0582	0.0874	0.0487	0.0208
soc-hamsterer	0.4476	0.5573	0.4136	0.4890	0.4483	0.5265	0.4016	0.4805	0.4567	0.5652	0.4068	0.4906
soc-pages-government	0.1699	0.1924	0.2295	0.1636	0.1680	0.1900	0.2392	0.1764	0.1749	0.1821	0.2331	0.1516
soc-pages-politician	0.1352	0.1429	0.1958	0.1225	0.1315	0.1314	0.2033	0.1258	0.1278	0.1268	0.1837	0.1106
soc-pages-shows	0.1094	0.1164	0.1808	0.1192	0.1262	0.1222	0.2122	0.1116	0.1094	0.1274	0.1838	0.0982
soc-wiki-elec	-0.0064	0.0609	0.0120	0.0358	-0.0036	0.0497	0.0079	0.0427	0.0063	0.0334	0.0106	0.0476
soc-advogato	0.3235	0.3516	0.2779	0.3425	0.3255	0.3522	0.2835	0.3432	0.3260	0.3403	0.2767	0.3453
soc-pages-public-figure	0.1880	0.1894	0.2480	0.1820	0.2010	0.1924	0.2586	0.1868	0.1986	0.1915	0.2473	0.1846
soc-anybeat	0.1470	0.2268	0.1045	0.0746	0.1431	0.2182	0.1275	0.0775	0.1449	0.2153	0.1064	0.0729
soc-pages-sport	0.1655	0.1482	0.1949	0.1383	0.1623	0.1485	0.1973	0.1481	0.1651	0.1488	0.1865	0.1466
soc-epinions	0.4176	0.5375	0.4658	0.5739	0.4232	0.5336	0.4650	0.5785	0.4167	0.5363	0.4685	0.5720
soc-pages-media	0.1666	0.1631	0.2204	0.1669	0.1689	0.1680	0.2239	0.1756	0.1731	0.1696	0.2200	0.1710
soc-pages-company	0.1400	0.1294	0.1836	0.1207	0.1397	0.1294	0.1898	0.1198	0.1455	0.1344	0.1833	0.1148
soc-gamesec-RO	0.1232	0.1093	0.1590	0.1010	0.1273	0.1142	0.1618	0.1026	0.1280	0.1116	0.1576	0.1023
Average correlation	0.1845	0.2157	0.2005	0.1889	0.1880	0.2104	0.2082	0.1930	0.1877	0.2130	0.1994	0.1873
Standard deviation	0.1227	0.1550	0.1230	0.1587	0.1233	0.1484	0.1191	0.1604	0.1229	0.1539	0.1224	0.1592

**Table 16 Fagin's intersection metric between the proposed and the traditional centrality measures and on considered real-world networks given in Table 7**

Instance name	HADC			EADC( $\alpha = 0.25$ )			EADC( $\alpha = 0.75$ )					
	BC	CC	EC	BC	CC	DC	EC	DC	CC	DC	EC	
	soc-gplus	0.3873	0.9108	0.6125	0.7472	0.3751	0.8643	0.6025	0.7626	0.3938	0.9309	0.6151
Twitter Lists	0.5074	0.9006	0.3708	0.2617	0.5237	0.7576	0.4133	0.2822	0.5052	0.9161	0.3659	0.2599
soc-hamsterer	0.6296	0.9318	0.7721	0.8624	0.6266	0.909	0.7797	0.8847	0.6317	0.9428	0.7679	0.8499
soc-pages-government	0.3962	0.8576	0.6462	0.6835	0.3802	0.7979	0.6706	0.7401	0.4024	0.8853	0.6307	0.6543
soc-pages-politician	0.4003	0.7377	0.5264	0.5629	0.3317	0.5951	0.6021	0.6447	0.4166	0.7894	0.4984	0.5246
soc-pages-shows	0.3818	0.8164	0.6155	0.7474	0.351	0.7154	0.6712	0.766	0.3912	0.8449	0.5841	0.7176
soc-wiki-elec	0.6856	0.9279	0.8281	0.9099	0.6869	0.92	0.8406	0.9223	0.6841	0.9296	0.8184	0.8993
soc-advogato	0.6855	0.9322	0.7789	0.8504	0.6866	0.9156	0.7907	0.867	0.6842	0.9396	0.7724	0.8416
soc-pages-public-figure	0.4086	0.4934	0.7131	0.563	0.2974	0.3305	0.7534	0.7474	0.4504	0.5715	0.6575	0.478
soc-anybeat	0.489	0.9049	0.7134	0.8538	0.4863	0.893	0.7156	0.863	0.4909	0.9122	0.7118	0.8467
soc-pages-sport	0.554	0.9609	0.5484	0.7228	0.5641	0.9344	0.5653	0.7328	0.551	0.9646	0.5438	0.7191
soc-epinions	0.4379	0.7928	0.7482	0.7065	0.4086	0.739	0.7541	0.7644	0.4508	0.8176	0.7389	0.6782
soc-pages-media	0.5299	0.8232	0.6674	0.8256	0.5213	0.7871	0.6852	0.8472	0.5327	0.8369	0.6599	0.8169
soc-pages-company	0.5619	0.95	0.5353	0.2957	0.5618	0.9123	0.5564	0.3019	0.5614	0.9566	0.531	0.2948
soc-gamesec-RO	0.5634	0.9371	0.4429	0.4533	0.567	0.8887	0.4712	0.4845	0.5634	0.9401	0.4402	0.4483
Average correlation	0.5079	0.8585	0.6346	0.6697	0.4912	0.7973	0.6581	0.7074	0.5140	0.8785	0.6224	0.6512
Standard deviation	0.1056	0.1203	0.1309	0.2032	0.1276	0.1617	0.1220	0.1998	0.0994	0.1007	0.1300	0.2049

## Conclusion

In this paper, we have proposed new hybrid centrality measures based on closeness (harmonic) and decay measures. The first hybridization is used to solve service coverage problem in flow networks where the demand for services is assumed to be proportional to the betweenness centrality. The proposed measures can also be used in another application that requires installing a facility farthest from the failure prone nodes. The solution in this case will be the node with minimum expected response-efficiency. The second hybridization attempts to find most ideal node for spreading information to the maximum fraction of population while considering a complex contagion. The proposed hybridization of centrality measures for both applications are based on the formulation for node-weighted centrality measures. The experimental results on several real-world networks show that the proposed measures perform relatively better than the individual traditional measures and also rank nodes differently.

Although, the reference centrality measures considered in this paper are due to the nature of considered applications, the framework allows using other measures based on the requirement. Analyzing the proposed measures on various other real-world networks and at the place of betweenness and degree centrality, hybridizing other measures specific to some particular applications are the possible future directions. The analysis of real-world networks where nonuniform weights are given at nodes using node-weighted centrality measures is another open direction. Another interesting direction is using hybrid centrality measures in multi-layer networks [47]. In multi-layer networks, we may use one layer of the network to compute node weights to be used for hybridization in other layers.

## Acknowledgements

The authors would like to thank the referees for comments that helped with the presentation of some of the material.

## Open Access

This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## Authors' contributions

AS, RRS and SRSI conceived the idea and contributed to the writing of the manuscript. RRS and AS worked on the implementation. All authors read and approved the final manuscript.

## Funding

RRS was funded in parts by the SERB-MATRICES Grant (MSC/2020/000374) and Research Initiation Grant from IIT Bhilai.

## Availability of data and materials

The data-sets analyzed during the current study are available in the repositories [31, 33–37, 39].

## Competing interests

The authors declare that they have no competing interests.

## Author details

<sup>1</sup> Department of Electrical Engineering and Computer Science, Indian Institute of Technology Bhilai, Raipur, Chhattisgarh, India. <sup>2</sup> Department of Computer Science and Engineering, Indian Institute of Technology, Ropar, Rupnagar, Punjab, India.

Received: 21 May 2020 Accepted: 30 October 2020

Published online: 13 November 2020

## References

1. Singh A, Singh RR, Iyengar S. Hybrid centrality measures for service coverage problem. In: International conference on computational data and social networks. Springer; 2019. p. 81–94.

2. Brandes U, Erlebach T. *Network Analysis: Methodological Foundations*, vol. 3418. Springer-Verlag, Berlin, Germany: Springer; 2005.
3. Jackson MO. *Social and economic networks*. Princeton: Princeton University Press; 2008.
4. Everett MG, Borgatti SP. The centrality of groups and classes. *J Math Sociol*. 1999;23(3):181–201.
5. Opsahl T, Agneessens F, Skvoretz J. Node centrality in weighted networks: generalizing degree and shortest paths. *Soc Netw*. 2010;32(3):245–51.
6. Abbasi A, Hossain L. Hybrid centrality measures for binary and weighted networks. In: *Complex networks*. Springer; 2013. p. 1–7.
7. Abbasi A. h-Type hybrid centrality measures for weighted networks. *Scientometrics*. 2013;96(2):633–40.
8. Wiedermann M, Donges JF, Heitzig J, Kurths J. Node-weighted interacting network measures improve the representation of real-world complex systems. *EPL (Europhysics Letters)*. 2013;102(2):28007.
9. Akanmu AA, Wang FZ, Yamoah FA. Clique structure and node-weighted centrality measures to predict distribution centre location in the supply chain management. In: *Science and information conference (SAI)*. IEEE; 2014. p. 100–111.
10. Freeman LC. Centrality in social networks conceptual clarification. *Soc Netw*. 1979;1(3):215–39.
11. Harary F. Status and contrastatus. *Sociometry*. 1959;22(1):23–43.
12. Boldi P, Vigna S. Axioms for centrality. *Internet Math*. 2014;10(3–4):222–62.
13. Anthonisse, J.M.: The rush in a directed graph. *Stichting Mathematisch Centrum. Mathematische Besliskunde (BN 9/71)*. 1971. p. 1–10.
14. Freeman LC. A set of measures of centrality based on betweenness. *Sociometry*. 1977;40(1):35–41.
15. Bonacich P. Factoring and weighting approaches to status scores and clique identification. *J Math Sociol*. 1972;2(1):113–20.
16. Puzis R, Elovici Y, Zilberman P, Dolev S, Brandes U. Topology manipulations for speeding betweenness centrality computation. *J Complex Netw*. 2014;3(1):84–112.
17. Zhang, X.J., Wang, Z.L., Zhang, Z.X.: Finding most vital node in satellite communication network. In: *Applied mechanics and materials*, vol 635. Trans Tech Publ; 2014. p. 1136–39.
18. Qiao, S., Peng, J., Li, H., Li, T., Liu, L., Li, H.: Webrank: a hybrid page scoring approach based on social network analysis. In: *Rough set and knowledge technology*. Springer; 2010. p. 475–82.
19. Lee GS, Djauhari MA. An overall centrality measure: the case of us stock market. *Int J Electr Comput Sci*. 2012;12(6):99–103.
20. Qiu L, Liang Y, Chen Z, Fan J. A new measurement for the importance of nodes in networks. In: *International Conference on Control Engineering and Information Systems*. CRC Press; 2014. p. 483–86.
21. Li-Qing Q, Yong-Quan L, Zhuo-Yan C. A novel algorithm for detecting local community structure based on hybrid centrality. *J Appl Sci*. 2014;14:3532–7.
22. Wang J, Rong L, Guo T. A new measure of node importance in complex networks with tunable parameters. In: *4th international conference on wireless communications, networking and mobile computing, WICOM'08*. IEEE; 2008. p. 1–4.
23. Buechel B, Buskens V. The dynamics of closeness and betweenness. *J Math Sociol*. 2013;37(3):159–91.
24. Kinney R, Crucitti P, Albert R, Latora V. Modeling cascading failures in the north American power grid. *Eur Phys J B Condens Matter Complex Syst*. 2005;46(1):101–7.
25. Buldyrev SV, Parshani R, Paul G, Stanley HE, Havlin S. Catastrophic cascade of failures in interdependent networks. *Nature*. 2010;464(7291):1025–8.
26. Motter AE, Lai Y-C. Cascade-based attacks on complex networks. *Phys Rev E*. 2002;66(6):065102.
27. Brandes U. A faster algorithm for betweenness centrality. *J Math Sociol*. 2001;25(2):163–77. <https://doi.org/10.1080/0022250X.2001.9990249>.
28. Singh RR, Goel K, Iyengar S, Gupta S. A faster algorithm to update betweenness centrality after node alteration. *Internet Math*. 2015;11(4–5):403–20.
29. Agarwal M, Singh RR, Chaudhary S, Iyengar, S. An efficient estimation of a node's betweenness. In: *Complex networks VI*. Springer; 2015. p. 111–21.
30. Singh RR, Iyengar S, Chaudhary S, Agarwal M. An efficient heuristic for betweenness estimation and ordering. *Soc Netw Anal Min*. 2018;8(1):66.
31. Rossi RA, Ahmed NK. The network data repository with interactive graph analytics and visualization. In: *Proceedings of the twenty-ninth AAAI conference on artificial intelligence*. 2015. <http://networkrepository.com>.
32. Brinkhoff T. A framework for generating network-based moving objects. *Geoinformatica*. 2002;6(2):153–80.
33. NRP-G-DATA. [https://www.iitk.ac.in/eeold/facilities/Research\\_labs/Power\\_System/NRP-G-DATA.pdf](https://www.iitk.ac.in/eeold/facilities/Research_labs/Power_System/NRP-G-DATA.pdf).
34. Barthelemy M. <https://www.quanturb.com/data.html>.
35. Batagelj V, Mrvar A. Pajek datasets. 2006. <http://vlado.fmf.uni-lj.si/pub/networks/data>.
36. Kunegis J. <http://konect.uni-koblenz.de/networks/>.
37. Leskovec J, Krevl A. SNAP datasets: Stanford large network dataset collection. 2014. <http://snap.stanford.edu/data>.
38. Son S-W, Kim H, Olave-Rojas D, Álvarez-Miranda E. Edge information of Chilean power grid with tap. *figshare*. Dataset. 2018. [https://figshare.com/articles/Edge\\_information\\_of\\_Chilean\\_power\\_grid\\_with\\_tap/6066587](https://figshare.com/articles/Edge_information_of_Chilean_power_grid_with_tap/6066587).
39. Watts DJ, Strogatz SH. Collective dynamics of 'small-world-networks'. *Nature*. 1998;393(6684):440–2.
40. Perna D, Interdonato R, Tagarelli A. Identifying users with alternate behaviors of lurking and active participation in multilayer social networks. *IEEE Trans Comput Soc Syst*. 2017;5(1):46–63.
41. Sprague DA, House T. Evidence for complex contagion models of social contagion from observational data. *PLoS One*. 2017;12(7):e0180802.
42. Centola D. *How behavior spreads: the science of complex contagions*, vol. 3. Princeton: Princeton University Press; 2018.
43. Calió A, Tagarelli A, Bonchi F. Cores matter? an analysis of graph decomposition effects on influence maximization problems. In: *12th ACM conference on web science*. 2020. p. 184–93.
44. Granovetter M. Threshold models of collective behavior. *Am J Sociol*. 1978;83(6):1420–43.

45. Shakarian P, Bhatnagar A, Aleali A, Shaabani E, Guo R. The independent cascade and linear threshold models. In: Diffusion in social networks. Springer; 2015. p. 35–48.
46. Weng L, Menczer F, Ahn Y-Y. Virality prediction and community structure in social networks. *Sci Rep*. 2013;3:2522.
47. Interdonato R, Magnani M, Perna D, Tagarelli A, Vega D. Multilayer network simplification: approaches, models and methods. *Comput Sci Rev*. 2020;36:100246.

### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

---

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)

---