IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received July 3, 2020, accepted July 12, 2020, date of publication July 15, 2020, date of current version July 24, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3009525

Interplay Between Hierarchy and Centrality in Complex Networks

STEPHANY RAJEH, MARINETTE SAVONNET, ERIC LECLERCQ, AND HOCINE CHERIFI, (Senior Member, IEEE)
Laboratoire d'Informatique de Bourgogne, University of Burgundy, 21000 Dijon, France
Corresponding author: Stephany Rajeh (stephany.rajeh@u-bourgogne.fr)

ABSTRACT Hierarchy and centrality are two popular notions used to characterize the importance of entities in complex systems. Indeed, many complex systems exhibit a natural hierarchical structure, and centrality is a fundamental characteristic allowing to identify key constituents. Several measures based on various aspects of network topology have been proposed in order to quantify these concepts. While numerous studies have investigated whether centrality measures convey redundant information, how centrality and hierarchy measures are related is still an open issue. In this paper, we investigate the association between centrality and hierarchy using several correlation and similarity evaluation measures. A series of experiments is performed in order to evaluate the combinations of 6 centrality measures with 4 hierarchy measures across 28 diverse real-world networks with varying topological characteristics. Results show that network density and transitivity play a key role in shaping the redundancy between centrality and hierarchy measures.

INDEX TERMS Hierarchy, centrality, complex networks, influential nodes.

I. INTRODUCTION

Networks offer a powerful representation of complex systems in which nodes represent the elementary units of the system and links represent their interactions. They are well adapted to investigate the relationships between structure and dynamics in such systems from the macroscopic to the microscopic level. At the microscopic level, the various roles of the nodes are induced by their specific pattern of connectivity. These interactions among nodes can cause them to exhibit varying levels of complexity. In turn, arising levels of complexity hinder the process of identifying the most influential nodes in a complex network [1]. Studying these interactions and quantifying the importance of the nodes have attracted a great number of researchers. Indeed, identifying key nodes is of prime interest in a wide range of strategic applications. Such applications span on fundamental fields such as prevention of epidemic spreading and vaccination strategies, prediction of genes using biomolecular interaction networks, maximizing information diffusion of marketing campaigns, ensuring robust power grids, identification of climate change and ocean currents, finding key heads in terrorist networks, and many more [2]. Although, characterizing the importance of the nodes in a network can be apprehended from various perspectives, the vast majority of works concern centrality and hierarchy.

Centrality is one of the main research topics in network science literature. Centrality measures quantify the ability of a node to influence the other nodes of the network based on the topological properties that are driving the network dynamics [3]. Early work has identified the key aspects of centrality. First, the number of local connections of a node and its ability to spread locally the information. Second, the position of a node in the network and its importance in the global exchange of information. A multitude of centrality measures have been proposed to date [4]. They can be classified into different groups. Indeed, they can be seen as global/path-based or local/neighborhood-based [5], [6]. They can be based on iterative-refinement process [6]. They can even be multidimensional by combining different measures together [7] or by incorporating the influence of the community structure [8]–[10].

Hierarchy is one of the foundations to understand complex networks, as they often take the form of hierarchical structure [11]. Indeed, hierarchy is ubiquitous throughout numerous natural or man-made complex networks. It can be observed in animal complex systems such as leader-follower network of pigeon flocks, biological complex networks such

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

as neural networks, transportation networks such as road networks, within organizations and between organizations, in social networks, and even the way we speak [12], [13]. One may ask why does hierarchy exist after all? Simon [11] tried to answer by his famous ''watchmaker parable''. It is having structure within systems such that each subsystem depends on another, and it results in more efficiency and resiliency than a random structure.

Several hierarchical measures based on different definitions of hierarchy have been developed [14]–[20]. Generally, they are used to quantify the hierarchical level of the whole network. Here, the focus is on the hierarchical level of individual nodes. One can distinguish two types of hierarchy measures in this case: nested and flow hierarchy. In nested hierarchy higher-level elements are contained in lower-level elements while in flow hierarchy nodes are arranged in different levels such that influential nodes are at a higher level and they are connected to the nodes they influence. Main nested hierarchy measures (k-core and k-truss) are based on hierarchical decomposition of nodes [19]–[21]. To our knowledge, Local Reaching Centrality'' (LRC) is the only flow hierarchy measure used to quantify node hierarchy [14].

Although the notions of hierarchy and centrality are quite different from a sociological point of view [22], the distinction seems more subtle when it comes to the associated measures. Indeed, both try to quantify the ''importance'' of a node based on topological information. One can consider that the most influential node is the one with the highest number of connections, or that connects two communities with each other, or the one allowing to reach all neighbors the fastest. This may be the case in some networks, but high centrality nodes aren't necessarily the most strategically located in terms of hierarchy [23]. In theory, these measures should capture different roles in the network. There have been some works on centrality measures incorporating some hierarchical aspects, such as social centrality, control centrality, and improved betweenness centrality [24]–[26].

Although there has been plenty of work [27]–[30] in order to study the interactions of the various centrality measures, this study is to our knowledge the first attempt to conduct a systematic empirical investigation about how centrality and hierarchy measures are related and what is the impact of the network topology on this relationship.

Centrality and hierarchy are two facets of influence. Therefore, whether the measures derived to quantify these two concepts reflect this duality is a fundamental issue. Indeed, despite conceptual differences between these two types of measures they may behave similarly on real-world networks. Moreover, it is of prime interest to know which measures can be associated in order to capture different aspects of network topology in a practical situation.

To investigate this issue, a set of six measures spanning the three main types of centrality (neighborhood-based, path-based and iterative refinement-based) together with a set of four hierarchy measures representing three categories (nestedness-based, flow-based, and mixed-based) are compared across a large set of real-world networks originating from various fields. Networks are chosen in order to span a wide range of basic topological property values such as density, transitivity and assortativity. A systematic evaluation of the various combinations of centrality and hierarchy measures is performed using three correlation measures (Pearson, Spearman, Kendall Tau and two similarity measures (Jaccard, RBO) for every network in order to explore to which extent the various combinations convey different information. Variations of these relations across networks are studied in order to clarify the relationship between hierarchy and centrality of nodes, and network topology.

The main contributions of this work are threefold:

1) A systematic investigation of how centrality measures are related to the main hierarchy measures in a number of real-world networks is performed.
2) The interplay between hierarchy and centrality measures is related to the basic network topological properties.
3) The most orthogonal hierarchy and centrality measures, whatever the network topological characteristics, are specified.

The rest of the paper is organized as follows. Section II introduces the hierarchy measures under study. Section III presents the centrality measures. The evaluation measures used to investigate the relationship between hierarchy and centrality are presented in section IV. Section V provides an overview of the datasets used and describes the methods applied. Section VI is devoted to the results and section VII develops the discussion. Finally, section VIII concludes the article.

II. HIERARCHY

We can distinguish two types of hierarchies when it comes to network structure:

i **Nested hierarchy**: Hierarchy imposed by a system considered on a higher level being composed of subsystems on a lower level that also are composed of subsystems, until we reach the lowest level (can also be called inclusive hierarchy) [31].

ii **Flow hierarchy**: Hierarchy imposed by the flow of resources that are essential for the system to be maintained. Entities on a higher level influence the ones at a lower level (can also be called control hierarchy) [14], [15], [31].

Flow hierarchy is related to the flow of resources essential to the network. It is typical of systems where resources are shared between entities to produce and maintain, such as food webs, software networks, and production and supply chain networks. Nested hierarchy is linked to the notion of organization through an inclusion process. The elementary parts of the system are divided into groups that are further divided into groups and so on until the core of the system is reached. The level of hierarchy is linked to the final group the parts belong to [32].

In network science literature, hierarchy is often used to quantify the macroscopic network organization [14]–[20]. In the following, we restrict our attention to the hierarchy measures linked to the microscopic network organization, i.e. at the node level. It can be defined as follows:

Hierarchy of nodes: Assume that $G(V, E)$ is an undirected and unweighted graph where V is the set of nodes of size $N = |V|$ nodes and $E \subseteq V \times V$ is the set of edges. The hierarchy of a node $v_i \in V$ is given by $\alpha(v_i)$. The function $\alpha(v_i)$ is a discrete measure when based on nestedness or mixed hierarchy ($\alpha(v_i) \rightarrow \mathbb{Z}^+$) and a real measure when based on flow hierarchy ($\alpha(v_i) \rightarrow \mathbb{R}^+$).

A. NESTED HIERARCHY

Nested hierarchy measures are based on the decomposition of the original network resulting in a number of nested entities such that each entity contains or is part of another. The network decomposition forms a hierarchy of induced subgraphs $G_k^f(V_k^f, E_k^f) \subseteq G(V, E)$ where f is the property characterizing the hierarchy and k is its value shared by the nodes.

Core and truss are the two main decomposition methods used to build nested hierarchy. For more information about hierarchical decomposition, one can refer to [21].

1) K-CORE NESTED HIERARCHY

The k-core of a graph G is the maximal subgraph G_k^c such that every node has at least k neighbors within the subgraph. The k-core subgraphs G_k^c are nested. Formally:

$$G \supseteq G_{k_{min}}^c \supseteq G_{k_{min}+1}^c \supseteq \ldots \supseteq G_{k_{min}+n}^c \ldots \supseteq G_{k_{max}}^c$$

where:

- k_{min} is the minimum degree value of the nodes in G
- $G_{k_{max}}^c$ is called the maximal k-core subgraph of G

A node v_i has a core number $c(v_i) = k$, if it belongs to a k-core G_k^c but not to the $(k+1)$-core G_{k+1}^c.

The k-core subgraphs can be extracted through a peeling process. Starting from the nodes with the minimum degree $k = k_{min}$, nodes that do not have a degree value of $k_{min}+1$ are removed from the graph and their core number is set to their degree value k_{min}. This process is iterated by computing the $k+j$-core. It ends when the maximum value k_{max} is reached such that the graph $G_{k_{max}}^c$ contains a non-empty k-core subgraph. The nested structure of k-cores reveals a hierarchy by containment. The level of hierarchy of a node v_i is given by:

$$\alpha_c(v_i) = k_{max} - (c(v_i) - 1) \tag{1}$$

Figure 1 represents a toy example to illustrate the hierarchical decomposition of a network according to the various hierarchy measures introduced in this section and their associated level of hierarchy. The k-core hierarchical decomposition of the given example is reported on the left side of the figure. The minimum degree of the nodes is one, so to extract the 1-core decomposition of the network G_1^c, nodes with a degree smaller than 2 are pruned. A core number $c(v_i) = 1$ is assigned to the set of removed nodes $V_1^c = \{2, 3, 4, 12, 14, 15, 16, 19\}$. Then all the nodes with a degree value smaller than 3 are removed from the remaining network G_2^c. The core number of the set of removed nodes $V_2^c = \{5, 6, 7, 8, 9, 10, 11, 13, 17, 26\}$ is set to $c(v_i) = 2$. The process continues by removing from G_3^c all the nodes with a degree value smaller than 4 and assigning to these nodes a core number $c(v_i) = 3$. Removing this set $V_3^c = \{1, 18, 20, 21, 22, 23, 24, 25\}$ of nodes, the maximum degree value of the original network is reached, and the remaining network G_4^c is empty.

The hierarchy level (α_c) of the nodes in the set V_k^c is given by $k_{max} - (j\text{-core} - 1)$. We have $k_{max} = 3$, therefore, V_3^c contains the nodes of hierarchy level 1 and so on.

2) K-TRUSS NESTED HIERARCHY

The truss decomposition is inspired by k-core. However, it considers the edges rather than the nodes, and the triangles they participate in. The k-truss of a graph G is the maximal subgraph G_k^t such that every link is contained in at least $k-2$ triangles within the subgraph. The k-truss subgraphs G_k^t are nested. Formally:

$$G \supseteq G_{k_{min}}^t \supseteq G_{k_{min}+1}^t \supseteq \ldots \supseteq G_{k_{min}+n}^t \ldots \supseteq G_{k_{max}}^t$$

where:

- k_{min} is the minimum number of triangles an edge engages in G such that $k_{min} \geq 2$
- $G_{k_{max}}^t$ is called the maximal k-truss subgraph of G

An edge e_{ij} has a truss number $t(e_{ij}) = k$, if it belongs to a k-truss G_k^c but not to the $(k+1)$-truss G_{k+1}^c.

The k-truss hierarchical decomposition can also be obtained through a peeling process. Starting from $k = k_{min}$, edges that are contained in at least k_{min} triangles but not in $k_{min} + 1$, are removed and their truss number is set to k_{min}. Note that k_{min} starts from 2. The set of resulting isolated nodes $V_{k_{min}-2}^t$ are also removed and assigned a truss value $k_{min} - 2$. The process is iterated by computing the $k_{min} + 1$-truss. It ends when the maximum value k_{max} such that the graph $G_{k_{max}}^t$ contains a non-empty k-truss subgraph is reached. k-truss also reveals a hierarchy by containment. The level of hierarchy of a node v_i is given by:

$$\alpha_t(v_i) = k_{max} - k_{min} - (t(v_i) - 1) \tag{2}$$

In the example of figure 1, the k-truss hierarchical decomposition is shown in the middle. Some edges are not engaging in any triangle, k_{min} is set to 2. This results in the 0-truss decomposition of the network G_0^t, where edges which do not participate in any triangle are removed. Accordingly, an edge truss number $t(e_{ij}) = 0$ is assigned to removed edges. The set of nodes isolated by the removal of the edges are also removed and assigned the same truss number $V_0^t = \{2, 3, 4, 12, 13, 14, 15, 16, 17, 19, 26\}$. Then, k is incremented to 3. All edges that engage in one triangle $(3 - k_{min})$ but not in two triangles are removed. They are assigned an edge trussness $t_e(e_{ij}) = 1$.

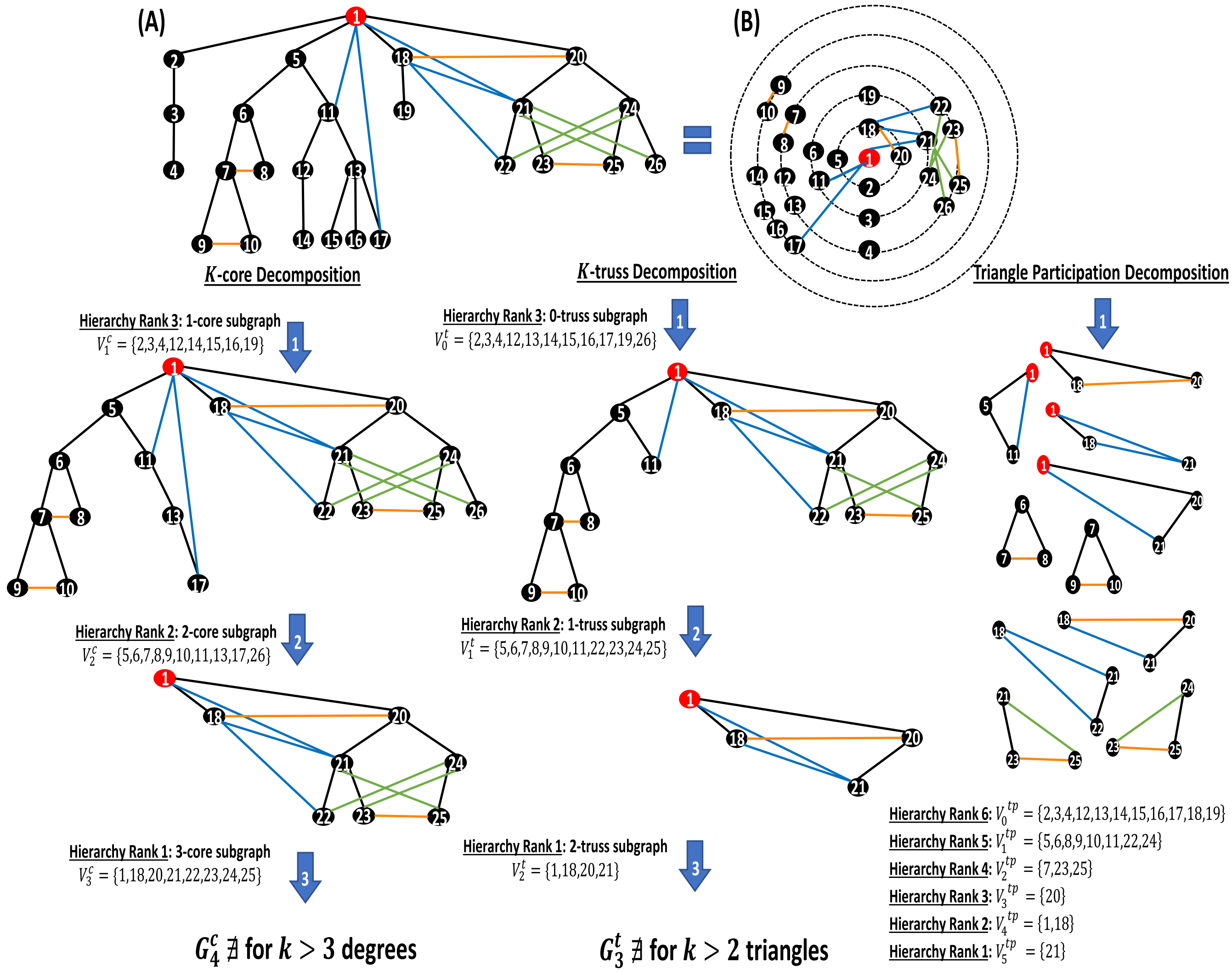


FIGURE 1. Obtaining hierarchy ranking of nodes for graph G with $N = 26$ and $E = 38$. (A) Example of tree-based view of G with irregularities and (B) radial-based view of G with irregularities. The irregularities are shown in different colors. The blue links indicate "shortcut" connections. The orange links indicate in-layer connections. The green links indicate fully connected trees at a specific rank. Hierarchical decomposition based on k-core (on the left) and k-truss (in the middle) and triangle participation (on the right) are illustrated.

The set of associated isolated nodes are also removed $V_1^t = \{5, 6, 7, 8, 9, 10, 11, 22, 23, 24, 25\}$ resulting in the 1-truss network G_1^t. The process continues by removing all edges that participate in two triangles and not three triangles are removed. The set of removed edges are assigned a truss number $t(e_{ij}) = 2$ together with the set of nodes isolated by this operation $V_2^t = \{1, 18, 20, 21\}$. The maximum trussness is reached at G_2^t because there is no edge participating in more than two triangles. The truss values range from 0 to 2.

The hierarchy level (α_t) of the nodes in the set V_k^t is given by $k_{max} - k_{min} - (j\text{-truss} - 1)$. We have $k_{min} = 2$, $k_{max} = 4$ and $t(v_i) = 2$, therefore, V_2^c contains the nodes of hierarchy level 1 and so on.

B. FLOW HIERARCHY

Many real-world networks such as information networks and production networks are better characterized by the flows of resources rather than a containment ordering. In such systems the entities are organized into a flow of hierarchy.

A hierarchy measure based on flow assigns a real hierarchy value to each node ($\alpha(v_i) \rightarrow \mathbb{R}^+$). The hierarchical level is based on this value. Indeed, if $|\alpha(v_i)| < |\alpha(v_j)| : \forall v_i, v_j \in V$ then node v_j is more important (higher hierarchically) than node v_i.

There have been a handful of work on flow hierarchy [15]–[17], but most of the work quantifies the hierarchy of the whole graph, which is not our focus. To our knowledge, Local Reaching Centrality (LRC) is one of the flow hierarchy measures quantifying the hierarchy of nodes.

1) LOCAL REACHING CENTRALITY FLOW HIERARCHY

LRC was originally developed for directed and weighted graphs, considered as a generalization of the m-reach centrality that takes into consideration nodes that are within m distance of a given node [14]. The authors consider $(m = N)$ where N is the number of nodes in G and define LRC as follows:

$$\alpha_l(v_i) = \frac{1}{(N-1)} \sum_{j=1}^{N} \left(\frac{\sum_{k=1}^{d^{out(v_i,v_j)}} \omega_{v_i}^{(k)}(v_j)}{d^{out}(v_i, v_j)} \right) \tag{3}$$

where:

- $d^{out}(v_i, v_j)$ is the length of the directed path that goes from node v_i to v_j via an outgoing edge such that $d^{out}(v_i, v_j) < \infty$
- $\omega_{v_i}^{k}$ is the weight of the k-th edge on its given path
- N is the total number of nodes in the network

However, LRC can be used for unweighted and undirected graphs. In this case, it reduces to closeness centrality for disconnected graphs:

$$\alpha_l(v_i) = \frac{1}{(N-1)} \sum_{j=1}^{N} \frac{1}{d(v_i, v_j)} \tag{4}$$

where:

- $d(v_i, v_j)$ is the distance between nodes v_i and v_j such that $d(v_i, v_j) < \infty$
- N is the total number of nodes in the network

From equation 3 and 4, we can see that the hierarchy is a continuous number between [0,1]. The higher the value, the more the node is able to impact other nodes, and the higher it is in terms of hierarchy.

C. MIXED HIERARCHY

We propose to consider "*Triangle Participation*" as a mixed hierarchy measure for nodes [22]. Indeed, it has the essence of both nested and flow hierarchy. Note that this measure is generally not used to quantify hierarchy but rather as a ratio in community detection problems [33]–[36]. This proposition is based on three main reasons:

i It can be considered as a flow hierarchy measure as it conveys flow of information probability according to how many times a node participates in a triangle. The higher the number of triangles' participation, the more the node is able to diffuse and acquire information.

ii Simultaneously, it can be viewed as a nested hierarchy measure as it allows to categorize the nodes according to the number of triangles they participate in. The hierarchy level of a node is therefore based on the density of the motifs that emerge as for k-core and k-truss.

iii k-core and k-truss are based on a more complex extraction process that may be prohibitive on large scale networks.

The hierarchy measure $(\alpha(v_i) \rightarrow \mathbb{Z}^+)$ divides the nodes V into groups $V_k^f \in G$ where f is the property function defined on the network and k is the property value shared by the nodes. Groups are not subgraphs since they are not based on higher levels containing lower levels as k-core and k-truss. If nodes v_i, v_j share the same property value k, then they share the same mixed hierarchy level $\alpha(v_i) \rightarrow \mathbb{Z}^+$. Hence, both nodes belong to the same group (set) where $v_i, v_j \in V_k^f$.

1) TRIANGLE PARTICIPATION MIXED HIERARCHY

To represent the mixed hierarchy measure, we need to first go on a couple of definitions. Considering graph $G(V, E)$:

- Let the adjacency matrix $A = (a_{i,j})$ describe the connectivity of the graph G such that:

$$a_{i,j} = \begin{cases} 1, & \text{if node } v_i \text{ is connected to node } v_j \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

- Let the neighborhood of any node v_i be defined as the set $\mathcal{N}_p(v_i) = \{v_j \in V : (v_i, v_j) \in E\}$ at length p, where $p = 1, 2, \ldots, D$. D is the diameter of G. Accordingly, two nodes are neighbors of order A^p if there's a minimal path connecting them at p steps.

Let nodes u, v, w be three nodes forming a closed triangle Δ_{uvw}. Triangle participation of a node v_i is simply the number of triangles it is in. It is defined as:

$$Tp(v_i) = \sum |\Delta_{uvw}| \tag{6}$$

where $|\Delta_{uvw}|$ is the number of times node v_i exists in triangle Δ_{uvw}, $\forall u, v, w \in V$. The level of hierarchy of a node v_i is given by:

$$\alpha_{tp}(v_i) = k_{max} + 1 - Tp(v_i) \tag{7}$$

Referring to the example in figure 1, the triangle participation hierarchical computation is shown on the right. Starting from $k_{min} = 0$, that is nodes not participating in any triangle, we obtain the first group of nodes $V_0^{tp} = \{2, 3, 4, 12, 13, 14, 15, 16, 17, 18, 19\}$. k_{min} is now incremented by 1, obtaining the second group of nodes, which participate in 1 triangle only $V_1^{tp} = \{5, 6, 8, 9, 10, 11, 22, 24\}$. Nodes in one group do not belong to another group as the k_{min} value is the highest number of triangles they are in. The process continues until we reach the highest number of triangles $k_{max} = 5$ with $V_5^{tp} = \{21\}$. As it can be seen, there is no a priori in this approach. The number of triangles for each node is counted only once (there's only one blue downward flash, unlike with k-core and k-truss decomposition). Moreover, there are more levels in mixed hierarchy. In other words, more heterogeneity among the nodes.

The hierarchy level (α_{tp}) of the nodes in the set V_k^{tp} is given by $k_{max} + 1 - Tp(v_i)$. There are $k_{max} = 5$ triangles, therefore, V_5^{tp} contains the nodes of hierarchy level 1 and so on.

III. CENTRALITY

Identifying the most influential nodes in a network using centrality measures is one of the main research issues in network science. For a complete overview about the various measures, the reader can refer to the surveys [2], [37]. The vast majority of centrality measures can be classified as neighborhood-based, path-based or iterative refinement-based measures [6]. While these measures are usually linked to a single topological property of the network, recent works turn to multidimensional definitions [7]–[10], [38]–[40].

Complexity is an important issue of centrality measurement. Measures can be classified as local or global. Global measures assume the knowledge of the overall network to compute the centrality of a node, while local measures need only information in the neighborhood of the node [5], [41]. Of course, global measures are generally more effective, but at the expense of a higher complexity that can be prohibitive, especially for large networks [42].

In order to explore the relations between centrality and hierarchy we choose to restrict our attention to the most influential centrality measures. According to the taxonomy adopted in [6], they belong to three main groupings: neighborhood-based, path-based, and iterative refinement-based centrality measures.

A. NEIGHBORHOOD-BASED CENTRALITY

Neighborhood-based centrality quantifies the importance of a node according to the influence it can exert on its local surroundings. Those local and semi-local measures are generally easy to compute, but they are totally agnostic about the overall network structure.

1) DEGREE CENTRALITY

The degree centrality of a node is one of the simplest centrality measures. It is proportional to the number of neighbors directly linked to the node. The more connections, the higher its influence on neighboring nodes. It requires low computation for identifying important nodes. It is defined as follows:

$$\beta_d(v_i) = \frac{1}{(N-1)} \sum_{j=1}^{N} a_{ij} \tag{8}$$

where:

- a_{ij} is obtained from A^1, 1-step neighborhood ($p = 1$)
- N is the total number of nodes in the network

2) LOCAL CENTRALITY

Local centrality extends degree centrality by increasing the size of the neighborhood of a node. While degree centrality consider direct neighbors, local centrality takes into consideration not only direct neighbors, but the neighbors of the neighbors. This is because the direct neighborhood alone may not be fully informative regarding a node's importance. It is defined as follows:

$$\beta_l(v_i) = \frac{1}{(N-1)} \sum_{j=1}^{N} a'_{ij} \tag{9}$$

where:

- a'_{ij} is obtained from A^2, 2-step neighborhood ($p = 2$)
- N is the total number of nodes in the network

B. PATH-BASED CENTRALITY

Path-based centrality measures quantify the ability of a node to spread information throughout the entire network. As they consider the paths going through each node, those global measures are difficult to compute in large-scale networks.

1) BETWEENNESS CENTRALITY

Betweenness centrality quantifies the importance of a node based on the fraction of shortest paths between any two nodes that pass through it. It is defined as follows:

$$\beta_b(v_i) = \frac{2}{(N-1)(N-2)} \left(\sum_{s,t \neq i} \frac{\sigma_{v_i}(v_s, v_t)}{\sigma(v_s, v_t)} \right) \tag{10}$$

where:

- $\sigma(v_s, v_t)$ is the number of shortest paths between nodes v_s and v_t
- $\sigma_{v_i}(v_s, v_t)$ is the number of shortest paths between nodes v_s and v_t that pass through node v_i
- N is the total number of nodes in the network

2) CURRENT-FLOW CLOSENESS CENTRALITY

Current-flow closeness centrality (also called information centrality) quantifies the node's importance according to the information transmitted along paths [43]. It is defined as follows:

$$\beta_c(v_i) = \frac{N}{\sum_{j=1} r_{ii} + r_{jj} - 2r_{ij}} \tag{11}$$

where:

- r_{ij} is the amount of information that can be transmitted from node v_i to v_j throughout all possible paths
- N is the total number of nodes in the network

r_{ij} is an element of the matrix R defined as follows: $R = [D - A + F]^{-1}$ where D is a $N-$dimensional diagonal matrix with the degree of the nodes along its diagonal and 0 everywhere else, A is the adjacency matrix of the network, and F is a matrix with all its elements equal to 1.

C. ITERATIVE REFINEMENT-BASED CENTRALITY

As path-based centrality measures, iterative refinement centrality measures make use of the topology of the overall network structure. Therefore, they are global centrality measures. However, in this case, a node's importance is linked to the importance of each of its neighbors.

1) KATZ CENTRALITY

Katz centrality quantifies the importance of a node in such a way that it takes into consideration the influence of all nodes and their paths with respect to it. However, as nodes become more distant from the node under study, their influence is attenuated. It is defined as follows:

$$\beta_k(v_i) = \sum_{p=1} \sum_{j=1} s^p a_{ij}^p \tag{12}$$

where:

- a_{ij}^p is the connectivity of node v_i with respect to all the other nodes at A^p
- s^p is the attenuation parameter where $s \in [0,1]$

As the distance between nodes p increases, the attenuation factor s^p decreases the influence of the other nodes v_j connected to v_i. Note that the attenuation parameter s should be strictly less than the inverse of the largest eigenvalue (λ_{max}) of the adjacency matrix A into have a solution for Katz centrality.

2) PageRank CENTRALITY

PageRank centrality works in a similar way that Katz centrality. It also takes into consideration the quantity and quality of nodes. However, PageRank is based on the probability of random walks. The adjacency matrix A is transformed to a stochastic matrix P representing the probability of visiting a node. Then, the centrality works iteratively using the power iteration till reaching a steady state. Originally, it is defined on directed graphs. The undirected version is defined as follows:

$$\beta_p(v_i) = \frac{1-d}{N} + d \sum_{v_j \in M_i} \frac{\beta_p(v_j)}{k_j} \tag{13}$$

where:

- $\beta_p(v_i)$ is the PageRank centrality of node v_i
- $\beta_p(v_j)$ is the PageRank centrality of node v_j
- M_i is the set of nodes linked to node v_i
- k_j is the number of links from node v_j to node v_i
- d is the damping parameter where $d \in [0,1]$ ensuring convergence in case $k_j = 0$
- N is the total number of nodes in the network

Note that the damping parameter value d is fixed at 0.85 in the subsequent experiments.

IV. EVALUATION MEASURES

In this section, the various correlation and similarity measures used to investigate the relationship between hierarchy and centrality measures are presented. In order to give more weight to our findings, we choose not to rely on a single evaluation measure but to use multiple ones. Indeed, doing so allows us to see if the various measures are in agreement. In addition to these measures, the k-means algorithm and the Schulze voting method used for deeper investigation are briefly presented.

A. CORRELATION

Three correlation measures are used to compare the set of hierarchy values with the set of centrality values for a given triplet (network, hierarchy measure, centrality measure). Pearson correlation is based on the values to compare, while Spearman correlation and Kendall Tau are rank-based comparisons.

1) PEARSON CORRELATION

Pearson's correlation coefficient is a popular measurement for the linear strength and direction between two variables. Its value ranges between $[-1, +1]$. The value -1 indicates a high negative correlation while the value $+1$ indicates a high positive correlation. A value of 0 means that there is no correlation at all.

Assume that the set of hierarchy measures α_i and that the set of centrality measures β_i for a network of size N is given, Pearson's correlation coefficient between the two measures is computed as follows:

$$\rho_p(\alpha, \beta) = \frac{\sum_{i=1}^{N} (\alpha_i - \bar{\alpha}) \left(\beta_i - \bar{\beta}\right)}{\sqrt{\sum_{i=1}^{N} (\alpha_i - \bar{\alpha})^2 \sum_{i=1}^{N} \left(\beta_i - \bar{\beta}\right)^2}} \tag{14}$$

where:

- $\bar{\alpha} = \frac{\sum_i^N \alpha_i}{N}$ is the average of the hierarchy measure values
- $\bar{\beta} = \frac{\sum_i^N \beta_i}{N}$ is the average of the centrality measure values
- N is the number of nodes in the network

2) SPEARMAN CORRELATION

Spearman's correlation coefficient is a modified version of Pearson's. Considering the ranks of the variables instead of their raw value, it measures their monotonic relationship. Monotonic relationships are less restrictive than linear relationships in case there is large variance but a relationship between variables still exists. Spearman's correlation values range also between $[-1, +1]$. Assume that the set of hierarchy measures α_i and that the set of centrality measures β_i for a network of size N is given, Spearman's correlation coefficient between the two measures is:

$$\rho_s(\alpha, \beta) = \frac{\sum_{i=1}^{N}(R(\alpha_i) - \overline{R(\alpha)})(R(\beta_i) - \overline{R(\beta)})}{\sqrt{\sum_{i=1}^{N}(R(\alpha_i) - \overline{R(\alpha)})^2 \sum_{i=1}^{N}(R(\beta_i) - \overline{R(\beta)})^2}} \tag{15}$$

where:

- $R(\alpha_i)$ and $R(\beta_i)$ represent the rank of the i-th hierarchy measure and centrality measures respectively
- $\overline{R(\alpha)}$ and $\overline{R(\beta)}$ are the average value of the ranks of the hierarchy and centrality measures respectively
- N is the number of nodes in the network

3) KENDALL TAU'S CORRELATION

Kendall Tau's correlation is a modified version of Spearman's. It also considers ranks and its values ranges between $[-1, 1]$. Furthermore, it takes into consideration the order of ranks as well. Suppose that $R(\alpha)$ and $R(\beta)$ are the ranking lists of the hierarchy measure and centrality measure, respectively. Kendall Tau's correlation determines the strength of the ordinal association based on the concordance (ordered in the same way) and discordance (ordered differently) between the pairs in the ranking lists $R(\alpha)$ and $R(\beta)$. A pair of nodes v_i and v_j is concordant if $R(\alpha(v_i)) > R(\alpha(v_j))$ and $R(\beta(v_i)) > R(\beta(v_j))$ or if $R(\alpha(v_i)) < R(\alpha(v_j))$ and $R(\beta(v_i)) < R(\beta(v_j))$. While the pair is discordant if $R(\alpha(v_i))) > R(\alpha(v_j))$ and $R(\beta(v_i)) < R(\beta(v_j))$ or if $R(\alpha(v_i)) < R(\alpha(v_j))$ and $R(\beta(v_i)) > R(\beta(v_j))$. Kendall Tau's version that takes into consideration ties as well is used. It is denoted by τ_b, where the pair is tied if $R(\alpha(v_i)) = R(\alpha(v_j))$ and/or $R(\beta(v_i)) = R(\beta(v_j))$. Kendall

Tau's correlation coefficient is defined as follows:

$$\tau_b(\alpha, \beta) = \frac{n_c - n_d}{\sqrt{(n_c + n_d + x)(n_c + n_d + y)}} \quad (16)$$

where:

- n_c is the number of concordant pairs
- n_d is the number of disconcordant pairs
- x is number of tied pairs on the hierarchy α variable
- y is number of tied pairs on the centrality β variable

B. SIMILARITY

Two similarity measures are used to compare the set of hierarchy values with the set of centrality values for a given triplet (network, hierarchy measure, centrality measure). Jaccard similarity is used to measure the proportion of common nodes in the top-k values in the two sets, while Rank-Biased Overlap (RBO) allows to compare the top-k values and also the entire sets.

1) JACCARD'S SIMILARITY

Jaccard's similarity index measures the similarity between two finite sets of data. Suppose that α and β are two sets containing the labels of a group of nodes extracted using a hierarchy measure and a centrality measure, respectively from a given network. The Jaccard index is defined as the size of the intersection divided by the size of the union of the sample sets. Formally it is given by:

$$J(\alpha, \beta) = \frac{|\alpha \cap \beta|}{|\alpha \cup \beta|} \quad (17)$$

Jaccard's similarity index values range between [0,1]. If there are no common nodes in the two sets $J = 0$, and $J = 1$ if all the members of the first set exist in the second set.

2) RANK-BIASED OVERLAP SIMILARITY

Rank-Biased Overlap (RBO) [44] measures the similarity of two ordered sets. It is based on Jaccard's similarity index but with indefinite sets at specific depth d. It allows giving more weight to differences at the top of the ranked sets than differences further down.

Assuming α and β are two infinite ordered sets of the hierarchy and centrality measures respectively, the RBO between those two sets is defined as follows:

$$RBO(\alpha, \beta) = (1 - p)\sum_{d=1}^{\infty} p^{(d-1)} \frac{|\alpha_d \cap \beta_d|}{d} \quad (18)$$

where:

- p is the probability of continuing to the next rank, while $(1 - p)$ is the probability of stopping at a specific rank
- d is the depth reached on sets α, β from position 1
- $|\alpha_d \cap \beta_d|/d$ is the proportion of the similarity overlap between hierarchy and centrality sets at depth d

RBO values range between [0,1]. There is no similarity between the two ranked sets if its value is null. A value of 1 indicates that the two ranked sets are identical.

C. K-MEANS CLUSTERING ALGORITHM

k-means is an unsupervised clustering algorithm that categorizes objects based on their feature values. The k parameter specifies the number of clusters, it is an input of the algorithm. Given a set of m samples $(m_1, m_2, \ldots, m_o)$ where each sample is composed of a d-dimensional feature vector, k-means divides the m observations into disjoint clusters $C = \{C_1, C_2, \ldots, C_k\}$ by minimizing the within-cluster criterion:

$$\sum_{i=1}^{o} \min_{\mu_j \in C}(||m_i - \mu_j||^2) \quad (19)$$

where:

- m_i is the d-dimensional feature vector of sample i
- μ_j is the mean of the samples m in cluster C

The k-means algorithm is used to cluster networks according to the various evaluation measures (correlation and similarity) across all possible combinations of the hierarchy and centrality measures used.

D. THE SCHULZE METHOD

The Schulze method [45] is a voting scheme that gives a single-winner or a sorted list of winners according to votes as indicated by user preferences. It transforms the lists of ordered preferences (ties being allowed) into a matrix of pairwise preferences Ω. From this matrix, another matrix is extracted, expressing the strength between pairwise preferences Υ. Strength extraction is defined on directed paths. The weakest link from all possible strongest paths between two candidates determines the strength of the winning candidate. The Schulze method is given in Algorithm 1. It uses a variant of the Floyd-Warshall algorithm to compute the strength of the strongest paths.

The Schulze voting scheme is used to rank the strength of occurrence among the hierarchy and centrality combinations. The ranking depends on the combinations' correlation and similarity magnitude across the networks used.

V. DATA AND METHODS

This section presents briefly the data used in the experiments and the experimental process. Note that unweighted and undirected networks are used. In case the original network is made of multiple components, only the largest connected component is retained.

A. DATA

To investigate extensively the relationship between hierarchy and centrality, 28 real-world networks originating from various domains such as social, biological, ecological, infrastructure networks have been selected. Their sizes range from tens to thousands of nodes. A brief description of these networks is provided. Table 1 reports their basic topological properties. Note that all the datasets are available online [46]–[49].

TABLE 1. Basic topological properties of real-world networks under investigation. N is the network size. $|E|$ is the number of edges. k_{min}, k_{max} are the minimum and maximum degree, respectively. $< k >$ is the average degree. $< d >$ is the average shortest path. ν is the density. ζ is the transitivity (also called global clustering coefficient). $k_{nn}(k)$ is the assortativity (also called degree correlation coefficient). γ_{max} is the maximum coreness. φ_{max} is the maximum trussness. * indicates the topological properties of the largest connected component of the network in case it is disconnected.

	N	$\|E\|$	k_{min}, k_{max}	$< k >$	$< d >$	ν	ζ	$k_{nn}(k)$	γ_{max}	φ_{max}
Animal Social Networks										
Mammals	28	235	[4, 23]	16.78	1.34	0.716	0.727	-0.004	14	12
Insects	113	4,550	[42, 109]	80.53	1.28	0.798	0.785	-0.030	60	45
Birds*	117	304	[1, 21]	5.19	4.36	0.577	0.472	0.062	8	9
Reptiles*	496	984	[1, 17]	3.96	8.08	0.008	0.419	0.342	8	9
Biological Networks										
Mouse Visual Cortex	193	214	[1, 31]	2.21	4.22	0.011	0.004	-0.844	2	3
E. coli Transcription*	329	456	[1, 72]	2.77	4.84	0.008	0.023	-0.263	3	3
Yeast Protein*	1,458	1,993	[1, 56]	2.73	6.77	0.001	0.051	-0.207	5	6
Human Protein*	2,217	6,418	[1, 314]	5.78	3.84	0.002	0.007	-0.331	10	5
Human Social Networks										
Zachary Karate Club	34	78	[1, 17]	4.58	2.36	0.139	0.255	-0.475	4	5
Madrid Train Bombings	64	243	[1, 29]	7.59	2.69	0.120	0.561	0.029	10	11
Physicians*	117	465	[2, 26]	7.94	2.58	0.068	0.174	-0.084	6	5
Adolescent Health	2,539	12,969	[1, 36]	10.21	4.55	0.002	0.141	0.231	7	7
Miscellaneous Networks										
Les Misérables	77	254	[1, 36]	6.59	2.63	0.086	0.498	-0.165	9	10
World Metal Trade*	80	875	[4, 77]	21.62	1.72	0.276	0.459	-0.391	14	14
Adjective Noun	112	425	[1, 49]	7.58	2.53	0.068	0.156	-0.129	6	5
Internet Autonomous Systems	6,474	12,572	[1, 1458]	3.88	3.68	0.0006	0.009	-0.181	12	10
Infrastructure Networks										
U.S. States	49	107	[1, 8]	4.36	4.16	0.090	0.406	0.233	3	3
U.S. Airports	500	2,980	[1, 145]	11.92	3.01	0.023	0.351	-0.267	29	27
EuroRoads*	1,039	1,305	[1, 10]	2.51	18.31	0.002	0.035	0.090	2	3
U.S. Power Grid	4,941	6,594	[1, 19]	2.66	18.98	0.0005	0.103	0.003	5	6
Collaboration Networks										
NetScience Collaboration	379	914	[1, 34]	4.82	6.04	0.012	0.430	-0.081	8	9
CS Ph.D. Collaboration*	1,025	1,043	[1, 46]	2.03	11.55	0.001	0.002	-0.253	2	3
GrQc Collaboration	4,158	13,422	[1, 81]	6.45	6.06	0.001	0.628	0.639	43	44
AstroPh Collaboration*	17,903	196,972	[1, 504]	22.004	4.19	0.001	0.317	0.201	56	57
Online Social Networks										
Retweets Copenhagen	761	1,029	[1, 37]	2.70	5.30	0.003	0.060	-0.099	4	4
Facebook Ego	4,039	88,234	[1, 1045]	43.69	3.69	0.010	0.519	0.063	115	97
Facebook Politician Pages	5,908	41,729	[1, 323]	14.12	4.69	0.002	0.301	0.018	31	26
PGP-based Social Network	10,680	24,316	[1, 205]	4.55	7.57	0.0004	0.378	0.238	31	27

1) ANIMAL SOCIAL NETWORKS

- **Mammals (mammalia-sheep-dominance):** The nodes represent sheep. They are connected if they have a dominance fight against each other. Interactions convey the need for superiority within a group [46].
- **Insects (insecta-ant-colony1-day01):** The nodes represent ants. They are connected in case they are enclosed within the same trapezoidal regions. Interactions convey group membership [46].
- **Birds (aves-weaver-social):** The nodes represent weaver birds. They are connected if they use the same bird nest for roosting or building within a year. Interactions convey social projection bipartite [46].
- **Reptiles (reptilia-tortoise-network-fi):** The nodes represent tortoises. They are connected if they use the same hole in the ground for refuge. Interactions convey social projection bipartite [46].

2) BIOLOGICAL NETWORKS

- **Mouse Visual Cortex (bn-mouse-visual-cortex-2):** Nodes are neurons in the visual cortex of the brain that is responsible for processing visual information. Interactions represent fiber tracts that connect one neuron to another [46].
- **E. coli Transcription:** Nodes are Escherichia coli bacteria regulating the conversion of DNA to RNA,

Algorithm 1 Schulze Voting Method

```
Require: C candidates list and Ω(i, j) matrix consisting of
    the number of voters who prefer candidate i over candidate j
  // Step 1: Initialization of Strength Matrix
  for i, j = 0 in C where i ≠ j do
    if Ω(i, j) > Ω(j, i) then
      ϒ(i, j) ← Ω(i, j)
    else
      ϒ(i, j) ← 0
    end if
  end for
  // Step 2: Strongest Paths Calculation
  for i, j = 0 in C where i ≠ j do
    for k = 0 in C do
      if i ≠ k and j ≠ k then
        ϒ(j, k) ← max[ϒ(j, k), min(ϒ(j, i), ϒ(i, k))]
      end if
    end for
  end for
  // Step 3: Ranking Final Set
  for i, j = 0 in C where i ≠ j do
    if ϒ(i, j) > ϒ(j, i) then
      L[i] ← L[i] + 1
    end if
    sort(L[i])
  end for
  return L
```

responding to various biological signals. Interactions represent transcriptions and regulations of genes [47].

- **Yeast Protein (bio-yeast-protein-inter):** A protein-protein interaction network where a protein is connected to another in case a direct interchange takes place. Interactions represent physical reactions [46].
- **Human Protein (maayan-figeys):** A protein-protein interaction of human cells, obtained from a first large-scale study on humans. Interactions represent physical reactions [49].

3) HUMAN SOCIAL NETWORKS

- **Zachary Karate Club:** Members of a karate club in a university are connected in case they interact with each other outside the club. Interactions represent friendship [47].
- **Madrid Train Bombings:** Nodes are terrorists of the Madrid train bombing on March 11, 2004. Interactions represent contact among two terrorists [49].
- **Physicians:** Nodes are physicians in U.S. towns. Interactions between two physicians represent trust. There is a link between two physicians if one of them asks for advice from another, wants to discuss a given topic, or is his friend [49].
- **Adolescent Health:** Nodes are students asked to list 5 of their best female friends and 5 of their best male friends. Interactions represent friendship ties [49].

4) MISCELLANEOUS NETWORKS

- **Les Misérables:** Actors in Victor Hugo's novel 'Les Misérables' connected if they appear in the same chapter of the novel. Interactions represent co-appearances [47].
- **World Metal Trade (world_trade):** World metal trade in 1994. Nodes represent countries involving heavy metal or high-technology metal product manufactures. Interactions represent trades from one country to another [48].
- **Adjective Noun (adjnoun):** Nodes are the most common occurring adjectives and nouns in "David Copperfield" novel of Charles Dickens. Interactions represent adjacent positions of any pair of words in the novel [46].
- **Internet Autonomous Systems (AS-20000102):** Nodes are autonomous systems (AS). Interactions represent connections for exchanging information between two AS. The network is a snapshot of the Internet on the 2^{nd} of January, 2000 [47].

5) INFRASTRUCTURE NETWORKS

- **U.S. States (contiguous-usa):** Nodes are the states of America. An edge represents border sharing between any two states. Hawaii and Alaska are excluded because they aren't adjacent to the rest of the states [46].
- **U.S. Airports:** Nodes represent airports in America. They are connected if there is a direct flight between two corresponding airports. Interactions represent a direct transport connection among the airports [49].
- **EuroRoads (inf-euroroad):** Nodes are European cities. Roads across the European continent are the links connecting cities that may be within the same country or not. Interactions represent a direct transport connection among cities [46].
- **U.S. Power Grid:** Nodes are either a generator, transformer, or substation in the western states of America. Interactions between nodes represent a power supply line [49].

6) COLLABORATION NETWORKS

- **NetScience Collaboration (ca-netscience):** Nodes are researchers in Network Science. Interactions represent co-authorship of scientific papers [46].
- **GrQc Collaboration (ca-GrQc):** Nodes are researchers co-authoring in General Relativity and Quantum Cosmology. Interactions represent co-authorship of scientific papers [46].
- **CS Ph.D. Collaboration (ca-CSphd):** Nodes are Ph.D. students and their supervisors specializing in Computer Science. Interactions represent passing scientific knowledge [46].
- **AstroPh Collaboration (ca-AstroPh):** Nodes are researchers co-authoring in Astrophysics, obtained from

e-print arXiv. Interactions represent co-authorship of scientific papers [46].

7) ONLINE SOCIAL NETWORKS

- **Facebook Ego Network (ego-facebook):** Nodes are users on Facebook, collected by a survey of participants using the Facebook application. Interactions represent online friendships [46].
- **Twitter Retweets Copenhagen (rt-twitter-copen):** Nodes are Twitter users with their retweets gathered in a United Nations conference in Copenhagen about climate change. Interactions represent retweets [46].
- **Facebook Politician Pages (fb-pages-politician):** Nodes represent politician pages on Facebook from different countries. Interactions represent mutual likes among the politicians' pages [46].
- **PGP-based Social Network:** Nodes are users of the web of trust, sharing information under the Pretty Good Privacy (PGP) algorithm. Interactions represent mutual secure information sharing among users [49].

B. METHODS

A series of experiments are conducted in order to characterize the relations between the set of hierarchy measures $A = \{\alpha_c, \alpha_t, \alpha_l, \alpha_{tp}\}$ and the set of centrality measures $B = \{\beta_d, \beta_l, \beta_b, \beta_c, \beta_k, \beta_p\}$ using a set of real-world networks $S = \{S_1, S_2, \ldots, S_{10}, \ldots, S_{28}\}$, a set of correlation evaluation measures $\rho = \{\rho_p, \rho_s, \tau_b\}$ and a set of similarity evaluation measures $Sim = \{J, RBO\}$. In the following study, for the four hierarchy measures α_i ($i = 1 \rightarrow 4$) used, notation is as following: α_c is α_1, α_t is α_2, α_l is α_3, and α_{tp} is α_4. For the six centrality measures β_j ($j = 1 \rightarrow 6$) used, notation is as following: β_d is β_1, β_l is β_2, β_b is β_3, β_c is β_4, β_k is β_5, and β_p is β_6.

1) COMPARING THE VARIOUS COMBINATIONS OF CENTRALITY AND HIERARCHY MEASURES FOR EACH NETWORK

In the first set of experiments, the aim is to compare the hierarchy measures to the centrality measures two-by-two for a given network. These experiments allow us to answer the main question of this study, that is, do hierarchy measures convey complementary information as compared to centrality measures given the topology of the network? Figure 2 illustrates the experimental process. In order to compare a hierarchy measure α_i to a centrality measure β_j for a given network S_i of size N, the sample set of hierarchy $\{\alpha_i(v_1), \alpha_i(v_2), \ldots, \alpha_i(v_l), \ldots, \alpha_i(v_n)\}$ and the sample set of the centrality $\{\beta_j(v_1), \beta_j(v_2), \ldots, \beta_j(v_l), \ldots, \beta_j(v_n)\}$ measures are computed for all the nodes of the network. Then a comparison measure $\gamma(\alpha, \beta)$ is computed. This process is performed for the 28 networks under investigation using all the centrality (Degree, Local, Current-Flow Closeness, Betweenness, Katz, and PageRank) and hierarchy (k-core, k-truss, LRC, and triangle participation) measures two-by-two. So, for each network, 24 combinations of hierarchy and

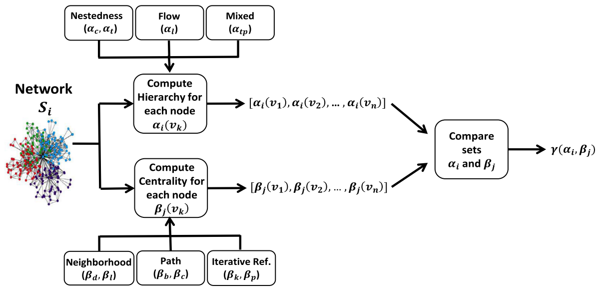


FIGURE 2. Comparing a hierarchy α_i and centrality β_j measure using an evaluation measure $\gamma(\alpha_i, \beta_j)$ for a given network S_i. k-core α_c and k-truss α_t are nested hierarchy measures. LRC α_l is a flow hierarchy measure. Triangle participation α_{tp} is a mixed hierarchy measure. Degree β_d, and Local β_l are neighborhood-based centrality measures. Betweenness β_b and Current-flow closeness β_c are path-based centrality measures. Katz β_k and PageRank β_p are iterative refinement-based centrality measures. The evaluation measure $\gamma(\alpha_i, \beta_j)$ can be a correlation or a similarity measure.

centrality measures are evaluated using Pearson, Spearman, and Kendall Tau correlation measures. Results of these experiments allow investigating if there are significant patterns that appear in terms of correlation. Additionally, the consistency of the pairwise correlation measures across networks can be evaluated. Finally, one can check if the results given by the various correlation measures are consistent.

Similarly, hierarchy measures are compared with centrality measures two-by-two using the two similarity measures (Jaccard and RBO) for a given network. Jaccard index checks the similarity across two finite sets regardless of rank. The sample set of centrality and the sample set of the hierarchy measures are computed for the top 10 nodes in small networks (N<150). For bigger networks, the top 10% are considered. RBO is designed for infinite sets taking into consideration the ranking of the nodes and their ties. Furthermore, it allows assigning a higher weight to top nodes using a tuning parameter p. Two values of the tuning parameter are used ($p = 0.5$ and $p = 0.9$). In order to compare with the results based on the Jaccard index, the similarity of the top 10 nodes for small networks or the top 10% for bigger networks are computed using RBO. Additionally, comparisons are performed considering all the nodes (the entire set of nodes) of the network for either small or big networks. Hence, there are 4 versions of RBO (at different p and on top-k and on the entire set of nodes). Results of these experiments allow checking the consistency across the similarity measures. Furthermore, comparisons can be performed with the results obtained using the correlation measures.

2) COMPARING THE NETWORKS ACCORDING TO THE EVALUATION MEASURES SAMPLE SETS

The second set of experiments are based on the previous results. The goal is to deepen the understanding of the interplay between centrality, hierarchy, and network topology. In other words, the following questions are raised: are there clusters of networks that can be discovered based on the correlation between centrality and hierarchy measures? That is, do networks show similar behavior based on their values of correlation and similarity between hierarchy and centrality combinations? And do the networks in these clusters share some specific topological properties? This investigation is based on three experiments.

In the first experiment, the goal is to check if the networks exhibit a similar behavior based on the various centrality and hierarchy combination. For a given evaluation measure γ, a network S_i is associated with the sample set $\Gamma_i = \{\gamma(\alpha_1, \beta_1), \gamma(\alpha_1, \beta_2), \ldots, \gamma(\alpha_i, \beta_j)\}$. For all the pairs of networks S_i, S_j, the Pearson correlation between their sample sets $\rho_p(\Gamma_i, \Gamma_j)$ is computed. Only Pearson correlation is used because, in this case, ranking is not our preference. This experiment is performed for all the evaluation measures under test. Figure 3 illustrates the experimental process. The corresponding correlation matrices are represented using a heatmap.

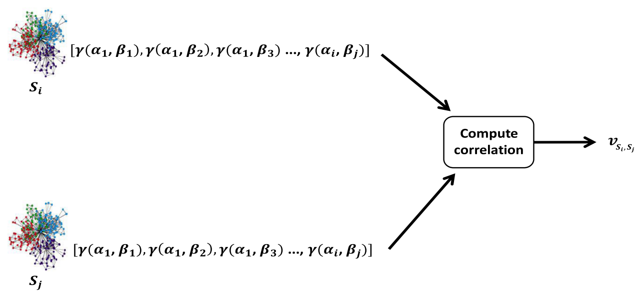


FIGURE 3. Comparing two networks S_i, S_j in terms of correlation after the computation of their hierarchy α_i and centrality β_j measure using an evaluation measure $\gamma(\alpha_i, \beta_j)$. a_i is a nested hierarchy measure (k-core α_c and k-truss α_t) or a flow hierarchy measure (LRC) α_l or a mixed hierarchy measure (Triangle participation α_{tp}). β_j is a neighborhood-based centrality measure (Degree β_d, and Local β_l) or a path-based centrality measure (Betweenness β_b and Current-flow closeness β_c) or an iterative refinement-based centrality measure (Katz β_k and PageRank β_p). The evaluation measure $\gamma(\alpha_i, \beta_j)$ can be a correlation or a similarity measure.

The second experiment aims to extract common topological characteristics from the networks based on their ranking. In this case, the evaluation measure between a centrality measure and a hierarchy measure is binarized. More precisely, $\gamma(\alpha_i, \beta_j)$ is set to one if the evaluation measure is above a threshold value μ and set to zero otherwise. Then, a score is assigned to each network according to the number of times the evaluation measure of the various combinations between centrality and hierarchy have been assigned a value of one. Finally, the networks are ranked according to this score.

Figure 4 illustrates this process. The same process is used for all the evaluation measures. In all the experiments the threshold value μ is set to 0.7. Indeed, it is generally admitted as a high value for similarity and correlation.

Finally, the third experiment aims at categorizing the sample set of networks S_i according to their multidimensional feature set $\Gamma_i = \{\gamma(\alpha_1, \beta_1), \gamma(\alpha_1, \beta_2), \ldots, \gamma(\alpha_i, \beta_j)\}$. The process based on the k-means algorithm is depicted in figure 5.

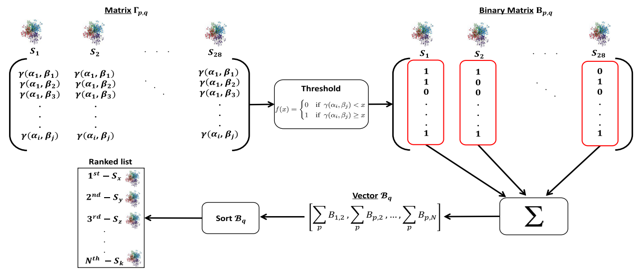


FIGURE 4. Ranking all networks $S = \{S_1, S_2, \ldots, S_{10}, \ldots, S_{28}\}$ in descending order after comparing the values of their evaluation measure $\gamma(\alpha_i, \beta_j)$ to a threshold $f(x)$. Starting from matrix $\Gamma_{p,q}$ a binary matrix $B_{p,q}$ is formed. Summing all the elements of its columns, the final ranking for each network is obtained. a_i is a nested hierarchy measure, and it includes k-core α_c and k-truss. α_t. LRC α_l as flow hierarchy measure. Triangle participation α_{tp} is a mixed hierarchy measure. β_j is a neighborhood-based centrality measure, and it includes Degree β_d, and Local centrality β_l. Betweenness β_b and Current-flow closeness β_c are path-based centrality measures. Katz β_k and PageRank β_p are iterative refinement-based centrality measures. The evaluation measure $\gamma(\alpha_i, \beta_j)$ can be either a correlation or a similarity measure.

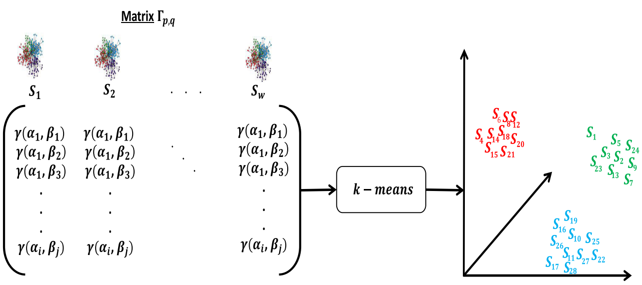


FIGURE 5. Clustering all networks $S = \{S_1, S_2, \ldots, S_{10}, \ldots, S_{28}\}$ using k-means algorithm according to their evaluation measure values $\gamma(\alpha_i, \beta_j)$ as features set. a_i is a nested hierarchy measure (k-core α_c and k-truss α_t) or a flow hierarchy measure (LRC α_l) or a mixed hierarchy measure (Triangle participation α_{tp}). β_j is a neighborhood-based centrality (Degree β_d and Local β_l) or a path-based centrality measure (Betweenness β_b and Current-flow closeness β_c) or an iterative refinement-based centrality measure (Katz β_k and PageRank β_p). $\gamma(\alpha_i, \beta_j)$ can be a correlation or a similarity measure.

3) COMPARING THE COMBINATIONS OF HIERARCHY AND CENTRALITY MEASURES USING THE SCHULZE VOTING METHOD

The third set of experiments aims to answer the question: what are the hierarchy measures which are the most distant from the centrality measures independently of the network topology? In order to explore this question, the Schulze voting method is used to rank the different combinations. The 28 networks S_i are considered as the voters. Given an evaluation measure γ, the 24 combinations between hierarchy and centrality measures are the candidates. Each network expresses a set of preferences about the 24 candidate combinations of centrality and hierarchy based on the magnitude of an evaluation measure $\Gamma_i = \{\gamma(\alpha_1, \beta_1), \gamma(\alpha_1, \beta_2), \ldots, \gamma(\alpha_i, \beta_j)\}$. The couple of centrality and hierarchy measures are then ranked according to the preference of all the voters. Note that, if a specific combination with high preference occurs frequently among the networks, it is highly ranked. Additionally, if a combination with lower preference also occurs frequently, it is also highly ranked, but at a lower rank that the one with higher preference and comparable frequency. The Schulze analysis is performed for all the correlation and similarity measures. A general view of the Schulze method using networks as voters is given in figure 6.

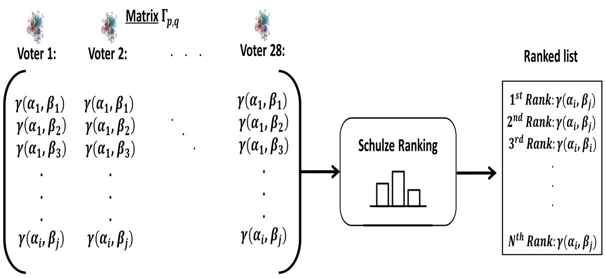


FIGURE 6. Ranking all possible combinations between hierarchy α_i and centrality β_j based on their evaluation measure values $\gamma(\alpha_i, \beta_j)$ using the Schulze Method. Networks are the voters and the evaluation of the combinations of the various centrality and hierarchy measures $\gamma(\alpha_i, \beta_j)$ are the candidates. α_i is a nested hierarchy measure (k-core α_c and k-truss α_t) or a flow hierarchy measure (LRC α_l) or a mixed hierarchy measure (Triangle participation α_{tp}). β_j is a neighborhood-based centrality measure (Degree β_d and Local β_l) or a path-based centrality measure (Betweenness β_b and Current-flow closeness β_c) or an iterative refinement-based centrality measure (Katz β_k and PageRank β_p). The evaluation measures $\gamma(\alpha_i, \beta_j)$ can be a correlation or a similarity measure.

VI. EXPERIMENTAL RESULTS

In this section, we report the results of the empirical evaluation about the relations between hierarchy, centrality and network topology based on the several experiments conducted. In order to improve reading fluency typical results are presented in this section, and complementary results are reported in supplementary materials.

A. COMPARING THE VARIOUS COMBINATIONS OF CENTRALITY AND HIERARCHY MEASURES FOR EACH NETWORK

In this series of experiments, the evaluation measures (3 correlation measures and 5 similarity measures) between the 6 centrality measures and the 4 hierarchy measures have been computed for each of the 28 networks. Heatmaps are used to present the results.

1) CORRELATION ANALYSIS

The correlation analysis is conducted using Pearson, Spearman, and Kendall Tau for the 24 combinations of the 6 centrality and the 4 hierarchy measures. Figure 7 reports the results for 6 networks illustrating the typical behavior of the 28 networks under investigation. The heatmap values range from the minimum correlation value (−0.007) observed in the entire dataset to 1. The color spectrum ranges from dark blue to yellow. In the following, three categories in terms of correlation range are considered. Low correlation in the range −0.007 to 0.4 is associated with blue colors in the heatmaps. Medium correlation values ranging from 0.4 to 0.8 are represented by green colors. Finally, high correlation values above 0.8 are colored in yellow. Let's look at Spearman correlation results presented in the left-hand side of figure 7. The heatmaps of 6 typical networks illustrating the results of this experiment are arranged from overall low correlation to overall high correlation.

The first typical behavior is illustrated by the heatmap of the CS Ph.D. collaboration network. It is also observed in the EuroRoads network. In this case, a large majority of correlation values are in the low and medium range. This translates into a heatmap where blue and dark green colors dominate. For these networks, we can conclude that there is no correlation between hierarchy and centrality measures. The second typical case is illustrated by the heatmap of E. coli Transcription network. It concerns also all the biological networks (Yeast Protein, Human protein, Mouse Visual Cortex), the U.S. Power Grids, and the Retweets Copenhagen networks. In this case, LRC and k-core are more or less well-correlated with the centrality measures. This translates into the heatmap with dominant colors ranging from light green to yellow. By contrast, low correlation values prevail for k-truss and triangle participation hierarchy measures. That is why dark green predominates in the heatmap.

The third case is illustrated by the heatmap of the Insects network. Mammals exhibit similar behavior. In this case, LRC and triangle participation are highly correlated with all the centrality measures. The dominant color of the heatmap is yellow for both hierarchy measures. k-core or k-truss show a quite different behavior with correlation values in the lower middle range. Indeed, dark green predominates in the heatmap.

The fourth case is shown in the heatmap of the Physicians network. It is also observed in the U.S. States and Adolescent Health networks. In this case, correlation values range from 0.53 to 0.94. Light green and yellow are the predominant colors of the heatmap. k-truss is the less correlated hierarchy measure, with its dark green color. It is followed by triangle participation and k-core with their light green colors. Finally, LRC, mostly yellow on the heatmap appears to be well-correlated with other centrality measures.

The fifth case is presented using the heatmap of the Birds network. GrQc and Reptiles networks share the same behavior. In this case, all the hierarchy measures exhibit high correlation values with the centrality measures. The dominant colors of the heatmap are light green and yellow. However, betweenness centrality does not correlate well with the hierarchy measures. Indeed, the predominant color for the betweenness line of the heatmap is dark green with correlation values ranging from 0.35 to 0.54.

The sixth and final case is represented by the heatmap of the World Metal Trade network. Zachary Karate Club, and Adjective Noun networks have quite similar heatmaps. In this case, almost all hierarchy and centrality measures are highly correlated. Dominant colors of the heatmaps are light green and yellow.

Turning to the heatmaps of Pearson and Kendall Tau's correlation, the results are presented respectively in the middle

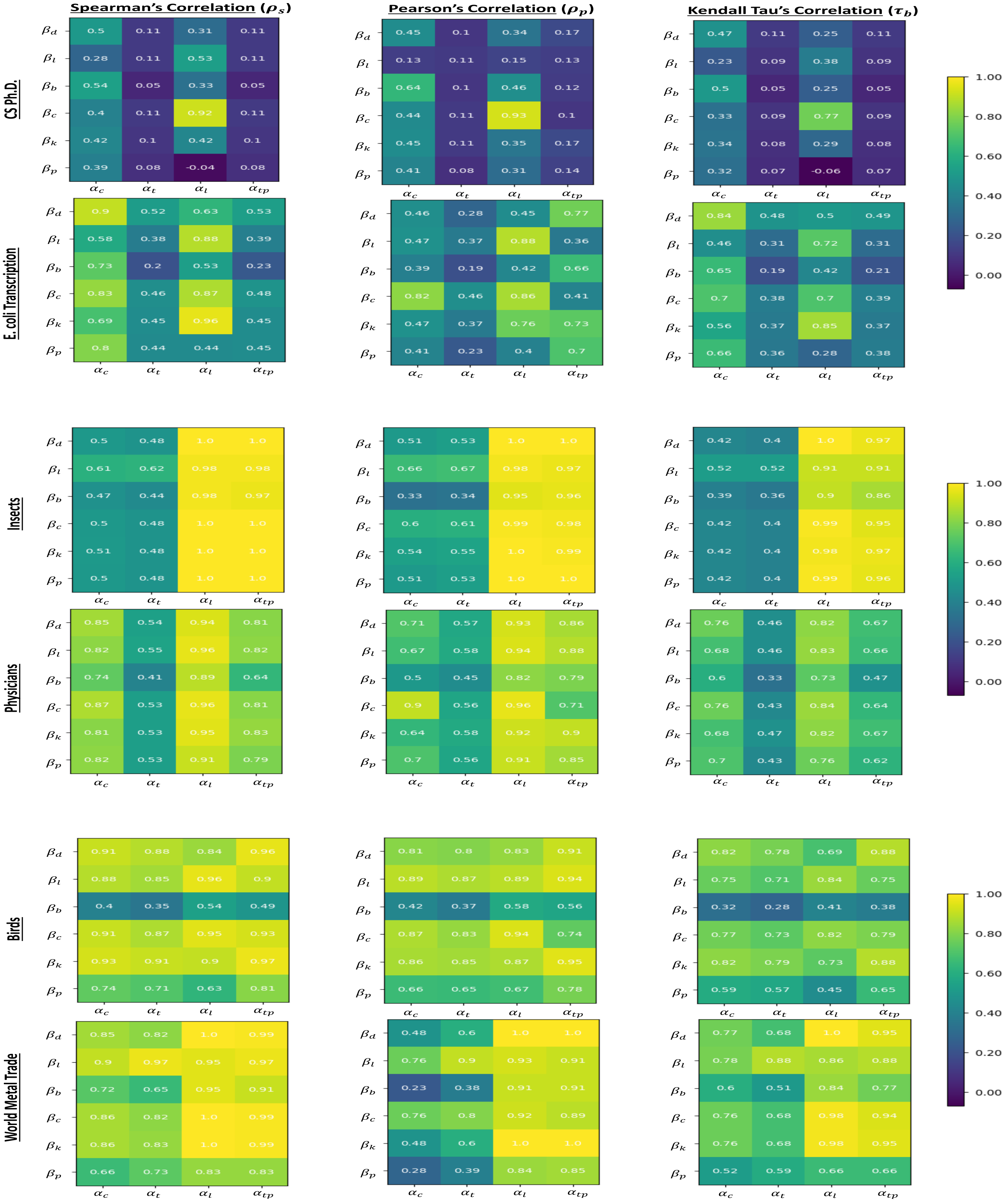


FIGURE 7. Heatmaps of the correlation evaluation measures for the various combinations of hierarchy α_i and centrality β_j measures of six real-world networks (from top to bottom). The hierarchy measures are $\alpha_c = k$-core, $\alpha_t = k$-truss, α_l = LRC, and α_{tp} = triangle participation. The centrality measures are β_d = Degree, β_l = Local, β_b = Betweenness, β_c = Current-flow Closeness, β_k = Katz, and β_p = PageRank. The correlation measure are Spearman (ρ_s), Pearson (ρ_p), and Kendall Tau (τ_b) from left to right.

and the right-hand columns of figure 7. Globally, it appears that Pearson correlation is closer to Spearman correlation values as compared to Kendall Tau. However, similar behavior is observed for both measures even if the correlation magnitudes are different. Note that the correlation magnitude of the Kendall Tau measure is systematically lower than Pearson and Spearman correlation corresponding values. This is not the case if Pearson is compared to Spearman. Indeed, Pearson correlation values can be higher or smaller than the corresponding Spearman correlation values. Consider for example the Insects network and the correlation between the degree centrality (β_d) and k-core (α_c) hierarchy measure. Spearman correlation is equal to 0.50. It is equal to 0.51 for Pearson and it decreases 0.42 for Kendall Tau.

Summing up the correlation analysis, one can say that 6 different correlation trends are observed between hierarchy and centrality measures across the 28 networks. Furthermore, the three correlation measures lead to quite consistent results.

2) SIMILARITY ANALYSIS

Results of similarity measures between the various hierarchy and centrality measures using the Jaccard index and RBO with its 4 versions have been computed. For the Jaccard index, top 10% nodes are used if the size of the network is $N \geq 150$ nodes while top 10 nodes are used if $N < 150$. RBO is computed on the same datasets for comparative purposes. Two values of the parameter p are used in RBO. A high value to give more weight to the top nodes ($p = 0.5$) and a medium value accounting for equal importance of the nodes ($p = 0.5$). Furthermore, RBO is computed using the top-k nodes and also the entire sample set.

Figure 8 reports the results for the 6 networks (from top to bottom) representative of the various behavior observed in the experiments and 3 out of 5 similarity measures (Jaccard, RBO top-k nodes with $p = 0.5$ and RBO top-k nodes with $p = 0.9$). The colors of the heatmaps vary from dark blue to yellow with similarity increasing from 0 to 1. The color range can be divided into 3 intervals. Blue indicates low similarity (0 to 0.4), green indicates medium similarity (0.4 to 0.8), and yellow indicates high similarity (0.8 to 1). Jaccard similarity is reported in the left-hand side of figure 8 to illustrate the various situations observed when evaluating the similarity of the 4 hierarchy measures with the 6 centrality measures. Networks are also arranged in increasing order between the two extreme cases (low similarity and high similarity).

E. coli Transcription is a typical example chosen to illustrate the first category. Reptiles, Yeast, EuroRoads, U.S. Power Grids, NetSci, GrQc, and AstroPh exhibit a quite similar behavior. In this case, low to medium similarity is observed among almost all hierarchy and centrality combinations. In the heatmap blue and dark green predominates.

The second category is represented by the Internet Autonomous Systems network heatmap. Mouse Visual Cortex, U.S. States and Retweets Copenhagen networks belong also to this category. In this case, similarity of LRC with centrality measures is high. It ranges from low to medium for the other hierarchy measures (k-core, k-truss, triangle participation). Except for LRC, blue and dark green predominate in the heatmap.

The third category is illustrated by the heatmap of Les Misérables. The heatmaps of Adjective Noun, PGP, and Facebook Politician Pages networks are quite comparable. In this case, the hierarchy measures show low similarity with a majority of centrality measures. However, LRC and triangle participation can exhibit high similarity with the few centrality measures. Blue and green predominate in the heatmap with few yellow patches.

The fourth category is represented by the heatmap of the Insects network. Mammals and Physicians networks are the other networks belonging to this category. In this case, k-core and k-truss have low similarity with all the centrality measures while LRC and triangle participation have high similarity. Blue and yellow predominate in the heatmaps.

The fifth category is illustrated by the Zachary Karate Club heatmap. It includes also Birds, World Metal Trade, Human Protein, and Facebook Ego networks. In this case, most combinations between centrality and hierarchy measures exhibit similarity in a medium range. Green predominates in the heatmaps.

The sixth category contains a single network, the U.S. Airports network. It shows significant similarity across almost all possible combinations between hierarchy and centrality measures. Yet, betweenness centrality departs from this behavior. In this heatmap light green to yellow predominates except for the betweenness line.

Results of the similarity evaluation using RBO top-k with $p = 0.5$ and $p = 0.9$ are reported in the middle and right-hand side of figure 8 respectively. The same networks illustrating the 6 categories uncovered using the Jaccard similarity measure are used for comparative purposes. Remember that Jaccard doesn't take into account the rank and ties to measure the similarity between two sets, while RBO does. Indeed, a high RBO value means that the two sets share a high proportion of common nodes with similar rankings. Globally, results are less mixed as compared to Jaccard similarity measure. The six categories observed while using Jaccard similarity measure reduce to 3 categories using RBO top-k with $p = 0.5$ (the middle column of figure 8).

The first category is characterized by a very low similarity of k-core and k-truss with all the centrality measures. In contrast, LRC and triangle participation exhibit a medium to high similarity with all the centrality measures except some rare exceptions. This category regroups a high number of the networks under study. It is illustrated in figure 8 by the heatmaps of E. coli Transcription, Internet Autonomous Systems, Insects, and U.S. Airports networks. Other networks that belong to this category are U.S. Power Grids, GrQc, Astroph, U.S. States, Retweets Copenhagen, Physicians, Birds, PGP, World Metal Trade, Madrid Train Bombings, and Mammals. Dark green predominates in the first two columns of the heatmap (k-core and k-truss), while light

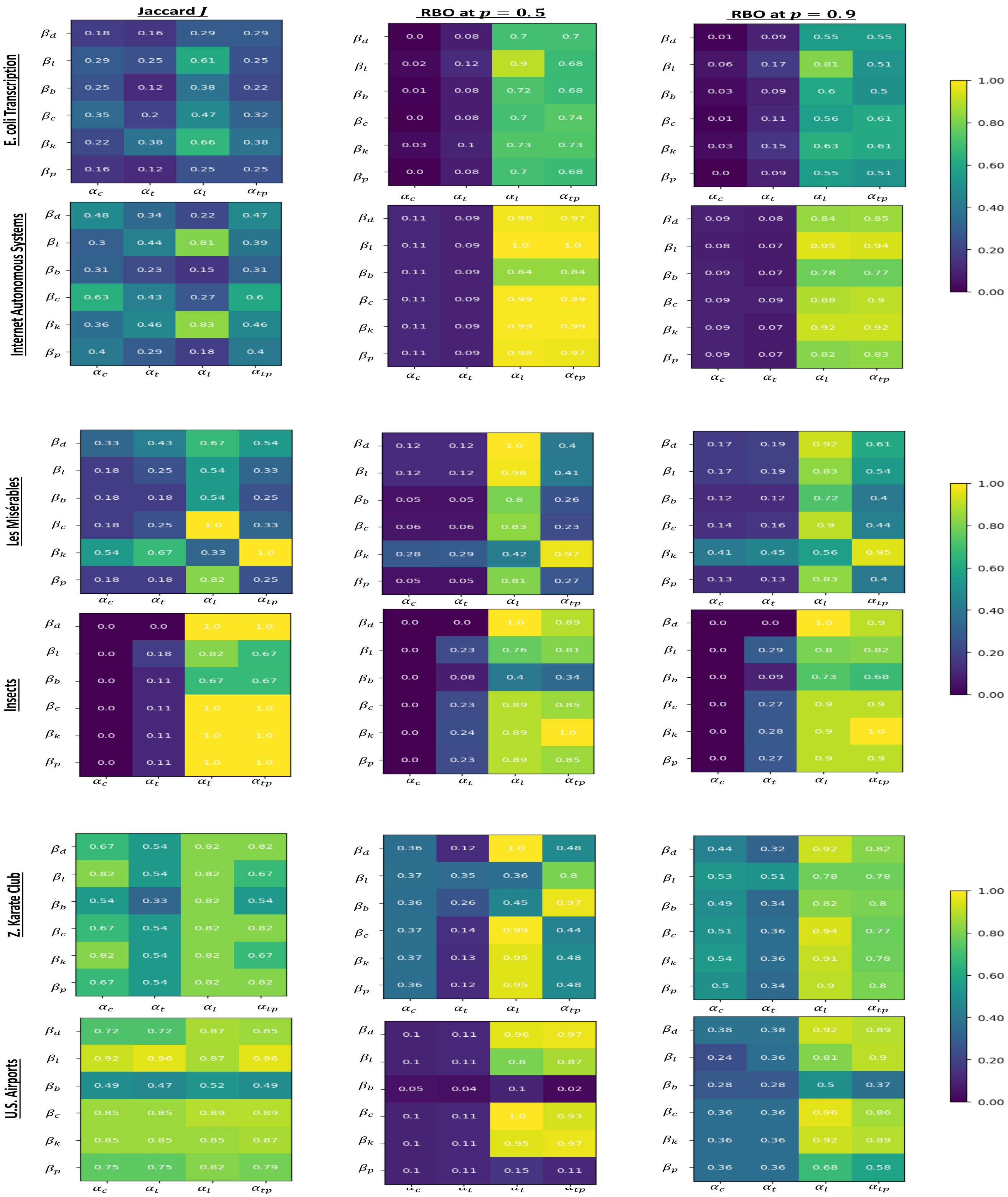


FIGURE 8. Heatmaps of the similarity evaluation measures for the various combinations of hierarchy α_i and centrality β_j measures of six real-world networks (from top to bottom). The hierarchy measures are $\alpha_c = k$-core, $\alpha_t = k$-truss, α_l = LRC, and α_{tp} = triangle participation. The centrality measures are β_d = Degree, β_l = Local, β_b = Betweenness, β_c = Current-flow Closeness, β_k = Katz, and β_p = PageRank. The similarity measures are Jaccard top-k nodes and RBO top-k nodes with $p = 0.5$ and $p = 0.9$ from left to right.

green and yellow predominate in the third and fourth column (LRC and triangle participation).

In the second category, LRC shows a high similarity with the centrality measures, while similarity with almost all centrality measures is low for k-core, k-truss, and triangle participation. Les Misérables, reported in figure 8, is a typical example of this behavior. Facebook Ego, Human Protein, and Mouse Visual Cortex are the other networks that belong to this category. Dark blue predominates in the heatmaps except on the third column (LRC).

In the third category, one can observe low to medium similarity among all hierarchy and centrality measures. It is illustrated in figure 9 by the heatmap of the Adolescent Health network. Other networks that fall into this category are Reptiles, Yeast Protein, EuroRoads, NetSci, CS Ph.D., and Facebook Politician Pages. Dark blue predominates the whole heatmap.

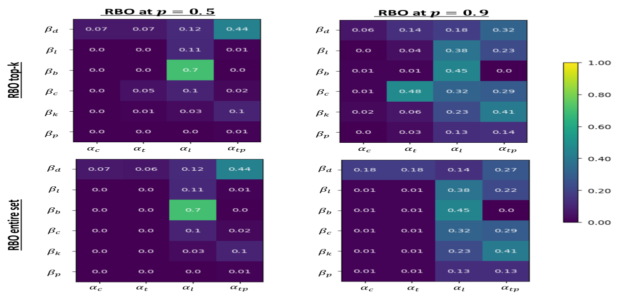


FIGURE 9. Heatmaps of RBO top-k and RBO of the entire set (top to bottom) with $p = 0.5$ and $p = 0.9$ (left to right) for the various combinations of hierarchy α_i and centrality β_j measures of the Adolescent Health Network. The hierarchy measures are $\alpha_c = k$-core, $\alpha_t = k$-truss, $\alpha_l =$ LRC, and $\alpha_{tp} =$ triangle participation. The centrality measures are $\beta_d =$ Degree, $\beta_l =$ Local, $\beta_b =$ Betweenness, $\beta_c =$ Current-flow Closeness, $\beta_k =$ Katz, and $\beta_p =$ PageRank.

Let's now compare RBO top-k at $p = 0.5$ to RBO top-k at $p = 0.9$. Globally the results are quite comparable. However, a closer look allows distinguishing three cases for a given hierarchy and centrality combination.

In the first case, increasing the p value increases the similarity value. Indeed, top nodes are given more importance and as they are identical and with the same rank, the similarity of the two sets increases. The combination of LRC hierarchy and PageRank centrality of the U.S. Airports network shown in figure 8 is a typical example. One can notice that $RBO_{p=0.5}(\alpha_l, \beta_p) = 0.15$ increases to $RBO_{p=0.9}(\alpha_l, \beta_p) = 0.68$.

In the second case, the similarity value decreases if p increases. Therefore, there are fewer top nodes identical in the two sets or their ranking differs. This results in a lower similarity for high p value. The combination between triangle participation hierarchy and degree centrality for the E. coli Transcription network shown in figure 8 illustrates this situation. Indeed, RBO decreases as p increases ($RBO_{p=0.5}(\alpha_{tp}, \beta_d) = 0.7$, $RBO_{p=0.9}(\alpha_{tp}, \beta_d) = 0.55$).

In the third case, similarity values at $p = 0.5$ and at $p = 0.9$ are identical. In this case, the overlap of nodes and their rankings do not significantly differ. Hence, increasing p does not have a notable effect on the similarity value. Such cases are observed for extreme values. An example of this case is the combination of k-core hierarchy and degree centrality in the Insects network in figure 8, where $RBO_{p=0.5}(\alpha_c, \beta_d) = 0$ whereas $RBO_{p=0.9}(\alpha_c, \beta_d) = 0$. Another extreme case in the same network concerns the similarity between LRC hierarchy and degree centrality, where $RBO_{p=0.5}(\alpha_l, \beta_d) = 1$ whereas $RBO_{p=0.9}(\alpha_l, \beta_d) = 1$. Note that the numbers are rounded so there may be a very small difference in the values. Yet, this is the characteristic of the third case, where we almost have the same similarity value regardless of the value of p.

Finally, let's compare RBO top-k nodes to RBO on the entire set of nodes. The results are quite similar. In other words, comparing the top-k nodes or all the nodes in the network does not provide much more information regarding the similarity of the hierarchy and centrality combinations. Figure 9 illustrates this observation with the Adolescent Health network. The complementary results of RBO using the entire dataset using the other networks are provided in supplementary materials.

Summing up the similarity analysis, using the Jaccard similarity one can classify the 28 networks used into 6 categories. This reduces to 3 categories when RBO is used. The first set of experiments does not allow getting a full picture of the relations between centrality and hierarchy measures. First of all, there is no clear relationship between correlation and similarity measures. One can notice that variations can be observed across networks. Nevertheless, some general trends seems to be emerging. Differences between the various centrality measures are not clear-cut, except for betweenness that sometimes behaves quite differently than the others. Another remark is that k-core and k-truss exhibit more often low similarity as compared to LRC and triangle participation. Finally, it appears that for all the correlation measures and the Jaccard similarity the results are more mixed as compared to RBO.

B. COMPARING THE NETWORKS ACCORDING TO THE EVALUATION MEASURES SAMPLE SETS

In order to relate the interactions between centrality and hierarchy measures to the network structure, we conduct three experiments based on the results of the previous one. In the first experiment, given an evaluation measure (correlation or similarity), the Pearson correlation is computed between the sample sets of the various combinations between centrality and hierarchy. The second experiment ranks the networks and the third one categorizes the networks based on the k-means algorithm.

1) CORRELATION ANALYSIS OF THE EVALUATION MEASURES SAMPLE SETS

For a given evaluation measure γ, a network S_i is characterized by the sample set of the evaluation measures' values between all the centrality and hierarchy measures

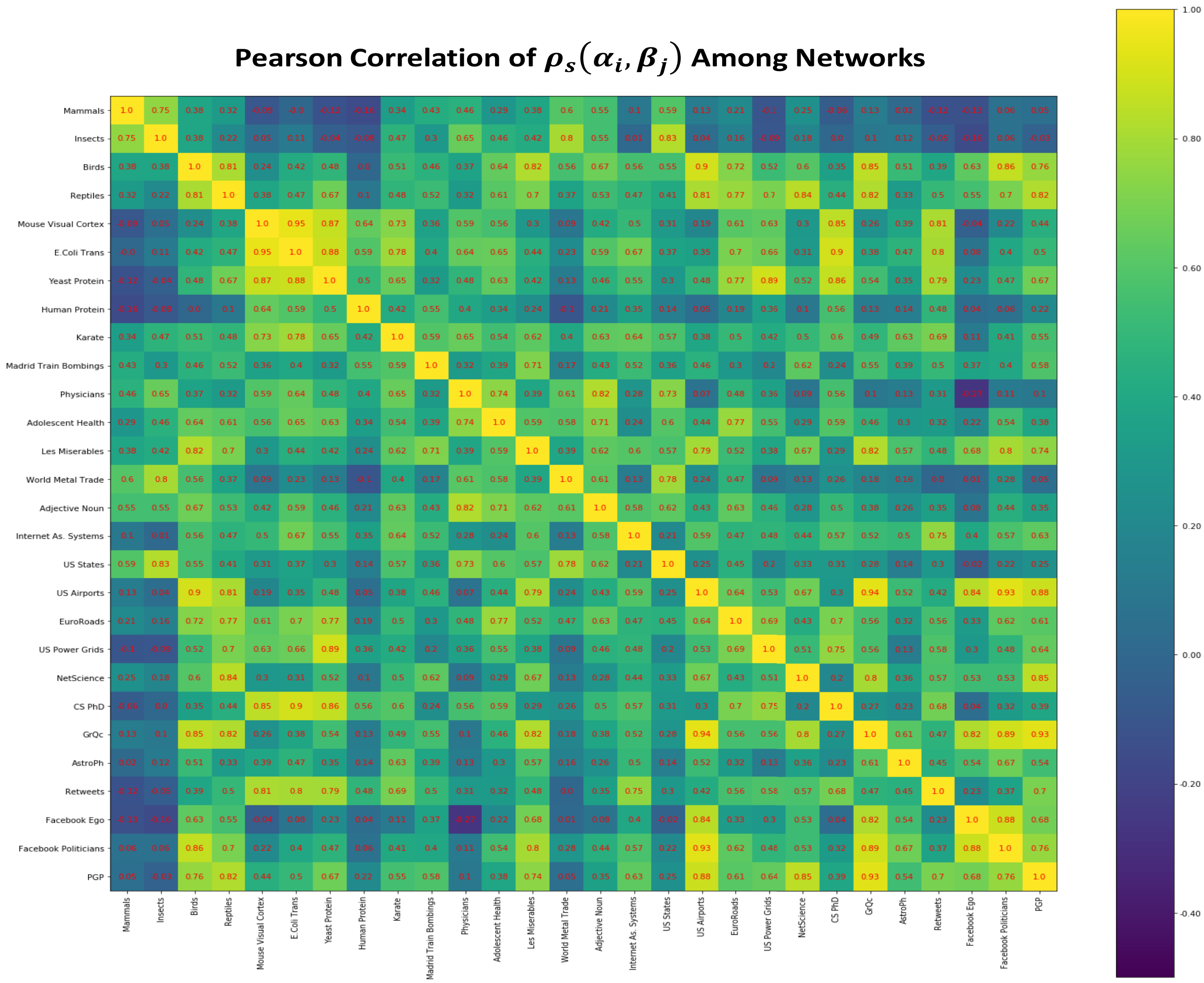


FIGURE 10. **Heatmap of the Pearson correlation between the network evaluation measures sample sets $\rho_p(\Gamma_i, \Gamma_j)$ for all the pairs of the 28 networks S_i, S_j. In the sample set $\Gamma_i = \{\gamma(\alpha_1, \beta_1), \gamma(\alpha_1, \beta_2), \ldots, \gamma(\alpha_i, \beta_j)\}$, the hierarchy measures are $\alpha_c = k$-core, $\alpha_t = k$-truss, $\alpha_l =$ LRC, and $\alpha_{tp} =$ triangle participation. The centrality measures are $\beta_d =$ Degree, $\beta_l =$ Local, $\beta_b =$ Betweenness, $\beta_c =$ Current-flow Closeness, $\beta_k =$ Katz, and $\beta_p =$ PageRank. The evaluation measure Γ is the Spearman correlation measure.**

$\Gamma_i = \{\gamma(\alpha_1, \beta_1), \gamma(\alpha_1, \beta_2), \ldots, \gamma(\alpha_i, \beta_j)\}$. In order to compare the networks two-by-two, the Pearson correlation between the sample sets $\rho_p(\Gamma_i, \Gamma_j)$ is computed. Experiments are performed using the three correlation measures and the two similarity measures. Results are reported for the Spearman correlation used as the evaluation measure in figure 10. One can observe that the heatmap is very patchy. It does not contain any clear pattern. Few yellow spots emerge now and then in a high proportion of green and blue. In other words, the vast majority of correlations are in the low and medium range. Nevertheless, some networks show a strong correlation. Results using the alternative correlation evaluation measures (Pearson and Kendall Tau) are reported in supplementary materials. Indeed, they are in the same vein and do not convey much more information.

Let's now turn to the similarity measures. Results using the Jaccard evaluation measure are reported in the supplementary material. Indeed, globally, they do not depart from the previous ones. However, the values of the correlation are much smaller and some patterns appear more clearly. For example, Facebook Ego, CS Ph.D., and Human Protein networks appear quite uncorrelated with a vast majority of the other networks. A less patchy matrix is observed using RBO as an evaluation measure. Especially when considering a high value of p ($p = 0.9$) as shown in figure 11. This confirms that the results are more consistent when RBO is used as an evaluation measure. The results for the other configurations of RBO are quite convergent. They are given in supplementary materials.

2) GROUPING THE NETWORKS BASED ON THE BINARIZED EVALUATION MEASURES SAMPLE SETS

In order to get a clearer picture about the relations between the network topology and the evaluation measures, the values of

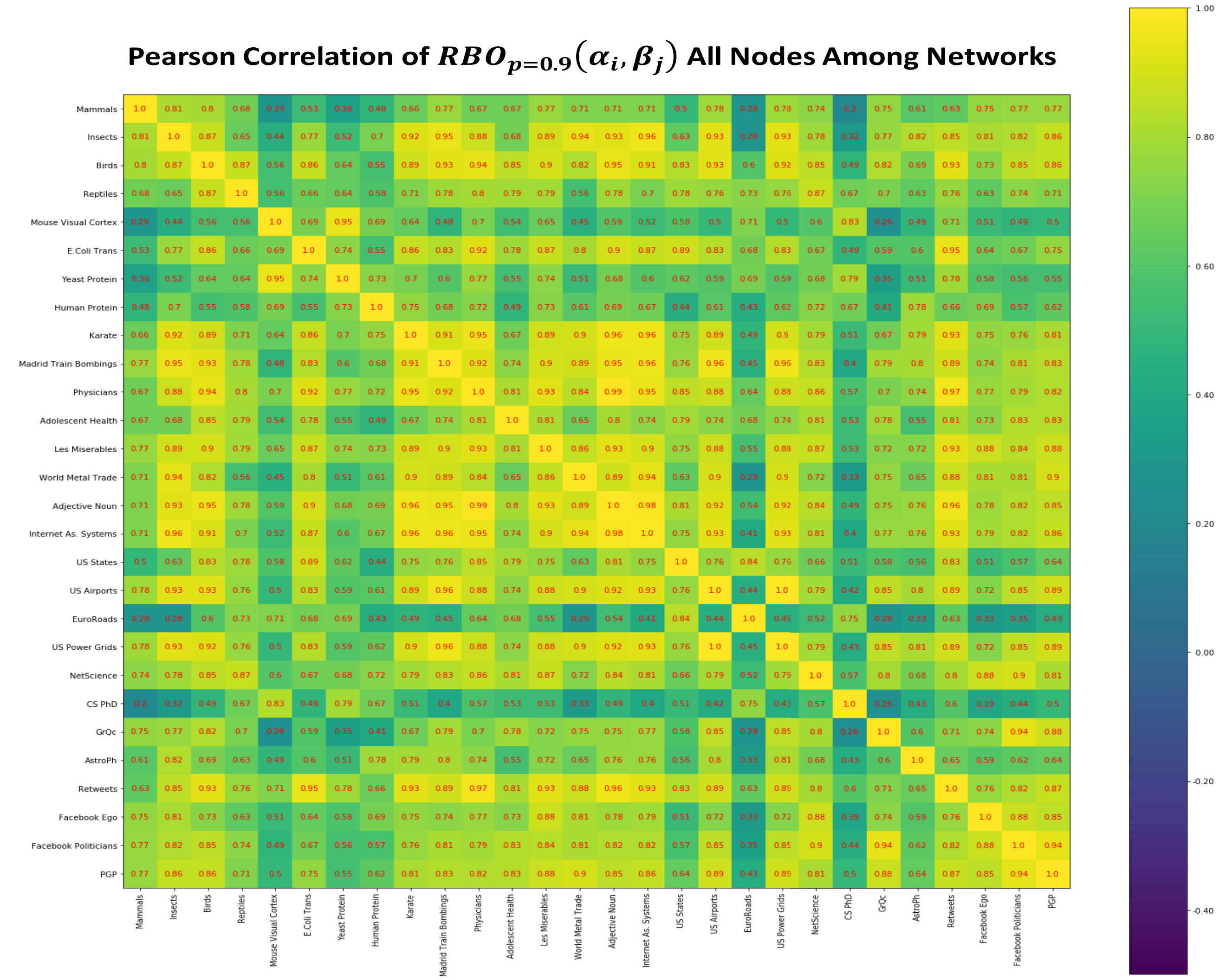


FIGURE 11. Heatmap of the Pearson correlation between the network evaluation measures sample sets $\rho_p(\Gamma_i, \Gamma_j)$ for all the pairs of the 28 networks S_i, S_j. In the sample set $\Gamma_i = \{\gamma(\alpha_1, \beta_1), \gamma(\alpha_1, \beta_2), \ldots, \gamma(\alpha_i, \beta_j)\}$, the hierarchy measures are α_c = k-core, α_t = k-truss, α_l = LRC, and α_{tp} = triangle participation. The centrality measures are β_d = Degree, β_l = Local, β_b = Betweenness, β_c = Current-flow Closeness, β_k = Katz, and β_p = PageRank. The evaluation measure Γ is the RBO similarity measure of the entire set of nodes with $p = 0.9$.

the evaluation measures $\gamma(\alpha_i, \beta_j)$ are binarized for each network. In other words, rather than using the continuous values of the evaluation measures, two cases are considered. Based on a threshold ($\mu = 0.7$), the evaluation measure for a given combination of centrality and hierarchy is set to 1 (meaningful) or 0 (not meaningful). Then the proportion of meaningful values out of the 24 combinations between centrality and hierarchy is computed for each network. Networks are then ranked according to the proportion of meaningful values, and topological properties of the networks are explored in order to check if an order relation is also uncovered. Finally, based on these relations the networks are categorized. This set of experiments is performed for both the correlation and the similarity measures.

Table 2 shows the networks ranked according to the meaningful values of the three correlation measures together with their basic topological properties (density (ν), transitivity (ζ) and assortativity ($k_{nn}(k)$)). The rankings based on each of the correlation measures (Spearman, Pearson, and Kendall Tau) are provided in the supplementary materials. Overall those values are well correlated (see table 3) and consequently, the rankings of the networks are quite similar except for few exceptions.

Looking at this table, we can clearly divide the networks into 3 groups based on their topological characteristics. The first group is made of the 8 networks with the higher ranks (From Adjective Noun to Mammals). One can notice that these networks exhibit in general high density (ν), high transitivity (ζ), and negative or near-zero assortativity ($k_{nn}(k)$). The second group contains 10 networks ranked in the middle range (Physicians to PGP). It is characterized in general by low density (ν), high transitivity (ζ), and positive assortativity ($k_{nn}(k)$). Finally, the third group is made of the 10 networks with the lowest ranks. Its typical features are low density (ν),

TABLE 2. Real-world networks sorted in decreasing order according to the proportion of meaningful correlation measures ($|\lambda_{\rho_s} + \lambda_{\rho_p} + \lambda_{\tau_b}| \geq 0.7$) denoted as λ_c. The network's basic topological properties are reported (ν is the density, ζ is the transitivity and $k_{nn}(k)$ is the assortativity).

Network	λ_c	ν	ζ	$k_{nn}(k)$
Adjective Noun	59/72	0.068	0.156	-0.129
Zachary Karate	55/72	0.139	0.255	-0.475
Les Misérables	55/72	0.086	0.498	-0.165
World Metal Trade	54/72	0.276	0.459	-0.391
U.S. Airports	52/72	0.023	0.351	-0.267
Madrid Train Bomb.	51/72	0.120	0.561	0.029
Birds	51/72	0.577	0.472	0.062
Mammals	49/72	0.716	0.727	-0.004
Physicians	41/72	0.068	0.174	-0.084
Facebook Pol. Pages	39/72	0.002	0.301	0.018
Facebook Ego	38/72	0.010	0.519	0.063
Insects	36/72	0.798	0.785	-0.030
U.S. States	34/72	0.090	0.406	0.233
AstroPh	34/72	0.001	0.317	0.201
GrQc	33/72	0.001	0.628	0.639
Adolescent Health	29/72	0.002	0.141	0.231
Reptiles	26/72	0.008	0.419	0.342
PGP	25/72	0.0004	0.378	0.238
Retweets Copenhagen	23/72	0.003	0.060	-0.099
Internet A. Systems	20/72	0.0006	0.009	-0.181
NetSci	19/72	0.012	0.430	-0.081
Human Protein	19/72	0.002	0.007	-0.331
E. coli Transcription	19/72	0.008	0.023	-0.263
Mouse Vis. Cortex	13/72	0.011	0.004	-0.844
Yeast Protein	12/72	0.001	0.051	-0.207
U.S. Power Grids	5/72	0.0005	0.103	0.003
EuroRoads	4/72	0.002	0.035	0.090
CS Ph.D.	3/72	0.001	0.002	-0.253

TABLE 3. Pearson correlation between the rankings of the networks based on their meaningful evaluation measures. Correlation measures are Spearman ($|\rho_s(\alpha, \beta)| \geq 0.7$) denoted as λ_{ρ_s}, Pearson ($|\rho_p(\alpha, \beta)| \geq 0.7$) denoted as λ_{ρ_p}, Kendall Tau ($|\tau_b(\alpha, \beta)| \geq 0.7$) denoted as λ_{τ_b}. Similarity measures are Jaccard ($|J(\alpha, \beta)| \geq 0.7$) denoted as λ_J, RBO at $p = 0.5$ ($|RBO_{p=0.5}(\alpha, \beta)| \geq 0.7$) denoted as $\lambda_{RBO}^{p=0.5}$, and finally RBO at $p = 0.9$ ($|RBO_{p=0.9}(\alpha, \beta)| \geq 0.7$) denoted as $\lambda_{RBO}^{p=0.9}$.

	λ_{ρ_s}	λ_{ρ_p}	λ_{τ_b}	λ_J	$\lambda_{RBO}^{p=0.5}$	$\lambda_{RBO}^{p=0.9}$
λ_{ρ_s}	1					
λ_{ρ_p}	0.84	1				
λ_{τ_b}	0.90	0.88	1			
λ_J	0.78	0.67	0.73	1		
$\lambda_{RBO}^{p=0.5}$	0.28	0.08	0.25	0.32	1	
$\lambda_{RBO}^{p=0.9}$	0.45	0.32	0.47	0.50	0.78	1

low transitivity (ζ), and negative or near-zero assortativity ($k_{nn}(k)$).

The process for relating the topological properties of the networks to their ranking according to the proportion of meaningful similarity values is slightly different than what has been done with the correlation evaluation measures. Table 3 reports the correlation between the various versions of RBO and Jaccard similarity measure. One can see that correlation is very low. Indeed, rankings according to Jaccard similarity are quite different than the rankings based on RBO. This is the reason why we prefer to consider Jaccard and RBO separately. Consequently, we do not aggregate the ranking of Jaccard and RBO to get an overall rank for each network.

Results for the ranked networks according to the proportion of meaningful Jaccard values together with their respective topological characteristics are given in table 4. One can see that the networks can be divided into 2 groups based on their topological characteristics. The first group of 7 top-ranked networks (from U.S. Airports to World Metal Trade) is characterized by high density (ν) and high transitivity (ζ), with negative or near-zero assortativity ($k_{nn}(k)$). The second group is made of the 21 remaining networks. They exhibit low or even zero proportion of meaningful Jaccard similarity, and are characterized by low density (ν). The other properties do not show clear trends. Indeed, transitivity (ζ) and assortativity ($k_{nn}(k)$) fluctuate in a wide range.

TABLE 4. Real-world networks sorted in decreasing order according to the proportion of meaningful Jaccard similarity measures $|J(\alpha, \beta)| \geq 0.7$ denoted as λ_J. The network's basic topological properties are reported (ν is the density, ζ is the transitivity and $k_{nn}(k)$ is the assortativity).

Network	λ_J	ν	ζ	$k_{nn}(k)$
U.S. Airports	20/24	0.023	0.351	-0.267
Zachary Karate	11/24	0.139	0.255	-0.475
Insects	9/24	0.798	0.785	-0.030
Madrid Train Bomb.	8/24	0.120	0.561	0.029
Birds	7/24	0.577	0.472	0.062
Mammals	5/24	0.716	0.727	-0.004
World Metal Trade	5/24	0.276	0.459	-0.391
Physicians	5/24	0.068	0.174	-0.084
Facebook Ego	4/24	0.010	0.519	0.063
Mouse Vis. Cortex	3/24	0.011	0.004	-0.844
Les Misérables	3/24	0.086	0.498	-0.165
Retweets Copenhagen	3/24	0.003	0.060	-0.099
Internet A. Systems	2/24	0.0006	0.009	-0.181
U.S. States	2/24	0.090	0.406	0.233
Human Protein	1/24	0.002	0.007	-0.331
CS Ph.D.	1/24	0.001	0.002	-0.253
Facebook Pol. Pages	1/24	0.002	0.301	0.018
PGP	1/24	0.0004	0.378	0.238
Reptiles	0/24	0.008	0.419	0.342
E. coli Transcription	0/24	0.008	0.023	-0.263
Yeast Protein	0/24	0.001	0.051	-0.207
Adolescent Health	0/24	0.002	0.141	0.231
Adjective Noun	0/24	0.068	0.156	-0.129
EuroRoads	0/24	0.002	0.035	0.090
U.S. Power Grids	0/24	0.0005	0.103	0.003
NetSci	0/24	0.012	0.430	-0.081
GrQc	0/24	0.001	0.628	0.639
AstroPh	0/24	0.001	0.317	0.201

For RBO, the proportion of meaningful similarities according to RBO top-k for the two p values ($RBO_{p=0.5}$ and $RBO_{p=0.9}$) are averaged. Indeed, both rankings are well-correlated. The detailed rankings of each evaluation measure considered individually are provided in the supplementary materials. The rankings of the networks with their topological properties are given in table 5. In this case, the networks can also be divided into 2 groups based on their topological

TABLE 5. Real-world networks sorted in decreasing order according to the proportion of meaningful RBO similarity ($|\lambda_{RBO_{p=0.5}} + \lambda_{RBO_{p=0.9}}| \geq 0.7$) denoted as λ_{RBO}. The network's basic topological properties are reported (ν is the density, ζ is the transitivity and $k_{nn}(k)$ is the assortativity).

Network	λ_{RBO}	ν	ζ	$k_{nn}(k)$
Adjective Noun	24/48	0.068	0.156	-0.129
Internet A. Systems	24/48	0.0006	0.009	-0.181
World Metal Trade	21/48	0.276	0.459	-0.391
Insects	21/48	0.798	0.785	-0.030
Madrid Train Bomb.	20/48	0.120	0.561	0.029
Physicians	19/48	0.068	0.174	-0.084
Zachary Karate	18/48	0.139	0.255	-0.475
U.S. Airports	16/48	0.023	0.351	-0.267
U.S. Power Grids	16/48	0.0005	0.103	0.003
Birds	15/48	0.577	0.472	0.062
AstroPh	14/48	0.001	0.317	0.201
Mammals	13/48	0.716	0.727	-0.004
Les Misérables	12/48	0.086	0.498	-0.165
U.S. States	12/48	0.090	0.406	0.233
Retweets Copenhagen	12/48	0.003	0.060	-0.099
PGP	12/48	0.0004	0.378	0.238
E. coli Transcription	10/48	0.008	0.023	-0.263
GrQc	9/48	0.001	0.628	0.639
Human Protein	7/48	0.002	0.007	-0.331
NetSci	5/48	0.012	0.430	-0.081
Facebook Ego	9/48	0.010	0.519	0.063
Mouse Vis. Cortex	3/48	0.011	0.004	-0.844
Yeast Protein	3/48	0.001	0.051	-0.207
Reptiles	2/48	0.008	0.419	0.342
CS Ph.D.	2/48	0.001	0.002	-0.253
Adolescent Health	1/48	0.002	0.141	0.231
Facebook Pol. Pages	1/48	0.002	0.301	0.018
EuroRoads	1/48	0.002	0.035	0.090

characteristics. The first group is composed of the 12 top-ranked networks (from Adjective Noun to Mammals). These networks exhibit high transitivity (ζ). Their density and transitivity values fluctuate in a wide range, with no particular visible trend. The 16 remaining networks forming the second category are characterized by low density (ν). No particular pattern is observed in their transitivity and assortativity values which vary in a wide range. Note that the rankings obtained with RBO using the entire node set provides very similar rankings to RBO top-k. That is the reason why they are presented in supplementary materials.

Table 6 summarizes the common properties shared by the various groups uncovered using the various type of evaluation measures. The density of the networks seem to be the most discriminating feature allowing to categorize the networks. Indeed, if one relies only on this parameter two categories emerge (high density, low density). This behavior is well-pronounced for correlation and Jaccard evaluation measures. The high density category made of the highly ranked networks is made of the same 6 networks out of 8 for correlation and 7 out of 8 for Jaccard. Comparing the 16 networks belonging to the low density category according to RBO to the corresponding category based on Jaccard and correlation evaluation measures, show that the results are

TABLE 6. Aggregation of basic topological characteristics of real-world networks after grouping them according to their meaningful correlation proportion from table 2, meaningful Jaccard similarity from table 4 and meaningful RBO similarity from table 5. The basic topological characteristics are: ν is the density, ζ is the transitivity, and $k_{nn}(k)$ is the assortativity. Two states can be given to the density and transitivity, either high denoted as H or low denoted as L. Two states can be given for assortativity, either positive denoted as P or negative denoted as N. If fluctuations occur within a group, they are denoted as X.

	ν	ζ	$k_{nn}(k)$
Correlation Rankings			
Group 1: Adjective Noun → Mammals	H	H	N
Group 2: Physicians → PGP	L	H	P
Group 3: Retweets Copenhagen → CS Ph.D.	L	L	N
Jaccard Rankings			
Group 1: U.S. Airports → World Metal Trade	H	H	N
Group 2: Physicians → AstroPh	L	X	X
RBO Rankings			
Group 1: Adjective Noun → Mammals	X	H	X
Group 2: Les Misérables → EuroRoads	L	X	X

also quite consistent. Fifteen out of sixteen belong to the low density category uncovered by the correlation evaluation measure and all of them belong to the low density category uncovered by the Jaccard evaluation measure. Transitivity is also an important feature influencing the classification. One can notice that, whatever the evaluation measure used, the category made of the top-ranked networks always group networks with high transitivity. Finally, the groupings are less sensitive to assortativity except for the correlation evaluation measures. It is worth noticing that whatever the evaluation measure, the networks belonging to the groups made of the top-ranked networks exhibit high transitivity. Furthermore, the networks belonging to the groups made of the low-ranked networks are sharing a low density. Hence, these two characteristics may be good predictors for two different groups. To sum up, these results allow us to conclude that the global topological characteristics of a network affect the relationship between hierarchy and centrality. Correlation is the most sensitive evaluation measures to highlight these relations, followed by Jaccard and RBO.

3) CLUSTERING THE NETWORKS BASED ON THE EVALUATION MEASURES SAMPLE SETS USING THE *K*-MEANS ALGORITHM

For a given evaluation measure γ, a network S_i is represented by a multidimensional vector $\Gamma_i = \{\gamma(\alpha_1, \beta_1), \gamma(\alpha_1, \beta_2), \ldots, \gamma(\alpha_i, \beta_j)\}$ made of the various combinations between the centrality and hierarchy measures. Therefore, these features can be used in order to categorize the networks using the k-means clustering algorithm. The main advantage as compared to the previous experiment is that no threshold is used to cluster the networks. Furthermore, comparisons can be performed with the results of the grouping based on the binarized evaluation measure. The main goal of these experiments is to check if the grouping made previously are consistent. This is why the value of the number of clusters k in the k-means algorithm is not based on an optimization criterion, but on the results of the previous experiment. Accordingly, k is set to 3 for the correlation evaluation measures, and $k = 2$ for the similarity evaluation measures (Jaccard and RBO).

Table 7 reports the content of the three clusters based on the Spearman correlation evaluation measure. It must be compared to the ranked networks based on the binarized evaluation measures reported in table 2. The first cluster regroups 10 networks. It has a big overlap with the first group based on the binarized evaluation measures that contains 8 networks. Both groups have 6 common networks (Zachary Karate Club, Madrid Train Bombings, Les Misérables, World Metal Trade, and Adjective Noun). The 4 remaining networks belong to the second group based on the binarized evaluation measures. The third cluster is made of 8 networks. Among these networks, 7 are common with the third group of table 2. The common networks are Mouse Visual Cortex, E. coli Transcription, Yeast Protein, EuroRoads, U.S. Power Grids, CS Ph.D., Retweets Copenhagen, and Human Protein. Globally, the three clusters uncovered by k-means have a high overlap with the grouping based on the ranking of the meaningful correlation evaluation measures. These results confirm our previous findings.

TABLE 7. Clusters of the networks S_i, S_j using k-means according to their Spearman correlation (ρ_s) values across all hierarchy and centrality combinations.

Cluster	Network
Cluster 1	Mammals, Insects, Zachary Karate Club, Madrid Train Bombings, Physicians, Adolescent Health, Les Misérables, World Metal Trade, Adjective Noun, U.S. States
Cluster 2	Birds, Reptiles, Internet Autonomous Systems, U.S. Airports, NetSci, GrQc, AstroPh, Facebook Ego, Facebook Politician Pages, PGP
Cluster 3	Mouse Visual Cortex, E. coli Transcription, Yeast Protein, Human Protein, EuroRoads, U.S. Power Grids, CS Ph.D., Retweets Copenhagen

Let's now turn to the comparison of the two clusters uncovered by the k-means algorithm with the two clusters of table 4. Cluster 1 in table 8 contains 7 networks while the corresponding cluster in table 4 contains 9 networks. They have 7 networks in common (Mammals, Insects, Birds, Zachary Karate Club, Madrid Train Bombing, World Metal Trade, and U.S. Airports). Those networks are characterized by high density and high transitivity. The second cluster in table 8 contains 19 networks while the corresponding group in table 4 contains 16 networks. They have 16 networks in common. This further illustrates the proximity of the two clustering methods.

Finally, let's compare the two k-means clusters reported in table 9 to the two groups of table 5. Those clusters are based on the $RBO_{p=0.9}$ of top-k nodes evaluation measures. There is a complete overlap between the clusters except for two networks which are AstroPh and Retweets Copenhagen.

TABLE 8. Clusters of the networks using k-means according to their Jaccard Similarity (J) values across all hierarchy and centrality combinations.

Cluster	Network
Cluster 1	Mammals, Insects, Birds, Zachary Karate Club, Madrid Train Bombings, Physicians, World Metal Trade, Adjective Noun, U.S. Airports
Cluster 2	Reptiles, Mouse Visual Cortex, E. coli Transcription, Yeast Protein, Human Protein, Adolescent Health, Les Miserables, Internet Autonomous Systems, U.S. States, EuroRoads, U.S. Power Grids, NetSci, CS Ph.D., GrQc, AstroPh, Retweets Copenhagen, Facebook Ego, Facebook Politician Pages, PGP

TABLE 9. Clusters of the networks using k-means according to their RBO top-k Similarity ($RBO_{p=0.9}$) values across all hierarchy and centrality combinations.

Cluster	Network
Cluster 1	Mammals, Insects, Birds, Zachary Karate, Madrid Train Bombings, Physicians, Les Misérables, World Metal Trade, Adjective Noun, Internet Autonomous Systems, U.S. States, U.S. Airports, U.S. Power Grids, Retweets Copenhagen
Cluster 2	Reptiles, Mouse Visual Cortex, E. coli Transcription, Yeast Protein, Human Protein, Adolescent Health, EuroRoads, NetSci, CS Ph.D., GrQc, AstroPh, Facebook Ego, Facebook Politician Pages, PGP

To summarize, clusters based on the k-means algorithm are very similar to the various groups identified by looking at the relations between the topological properties of the networks and the rankings based on the proportion of meaningful evaluation measures. These results strengthen our understanding of the relationship between the topology of networks and the interaction of centrality and hierarchy measures.

C. COMPARING THE COMBINATIONS OF HIERARCHY AND CENTRALITY MEASURES USING THE SCHULZE VOTING METHOD

In this section, the Schulze voting method is used in order to rank the output of the evaluations-measures for the various combinations of hierarchy and centrality measures. For a given evaluation measure γ, each of the 28 networks under test is considered as a voter S_i, and the evaluation measure $\gamma(\alpha_i, \beta_j)$ of hierarchy α_i and centrality measures β_j are the candidates. Voters rank the candidates in the descending order of the value of each of the 24 evaluation measures $\Gamma_i = \{\gamma(\alpha_1, \beta_1), \gamma(\alpha_1, \beta_2), \ldots, \gamma(\alpha_i, \beta_j)\}$. As voters rank the candidates in the order from the one that they most want to win to the one they least want to win, the outcome of the voting process is the set of evaluation measures ranked from the most correlated/similar to the least correlated/similar according to the population of voters. This experiment is performed using the three correlation measures (Pearson, Spearman, Kendall Tau) and the 5 similarity measures (Jaccard, RBO top-k at $p = 0.5$ and $p = 0.9$ and RBO of the entire set of nodes at $p = 0.5$ and $p = 0.9$).

Table 11 reports the rankings based on the 3 correlation measures. The Kendall Tau correlation between the ranking

TABLE 10. Kendall Tau correlation between the rankings of hierarchy and centrality combinations according to the Schulze Method for the various evaluation measures. Correlation evaluation measures are Spearman (ρ_s), Pearson (ρ_p), and Kendall Tau (τ_b). Similarity evaluation measures are Jaccard (J), RBO top-k at $p = 0.5$ and $p = 0.9$ ($R^{top}_{0.5}$ and $R^{top}_{0.9}$), RBO of the entire set of nodes at $p = 0.5$ and $p = 0.9$ ($R^{all}_{0.5}$ and $R^{all}_{0.9}$). The 28 real-world Networks are considered as voters.

	ρ_s	ρ_p	τ_b	J	$R^{top}_{0.5}$	$R^{top}_{0.9}$	$R^{all}_{0.5}$	$R^{all}_{0.9}$
ρ_s	1							
ρ_p	0.63	1						
τ_b	0.88	0.60	1					
J	0.40	0.55	0.36	1				
$R^{top}_{0.5}$	0.25	0.42	0.19	0.81	1			
$R^{top}_{0.9}$	0.25	0.41	0.19	0.80	0.92	1		
$R^{all}_{0.5}$	0.35	0.52	0.31	0.76	0.78	0.80	1	
$R^{all}_{0.9}$	0.34	0.51	0.28	0.74	0.76	0.79	0.91	1

TABLE 11. Ranking of the Schulze Method for all possible combinations between hierarchy α and centrality measures β. The ranks are sorted based on the average (SC_c) of all correlation evaluation measures γ used: Spearman (ρ_s), Pearson (ρ_p), and Kendall Tau (τ_b). The 28 real-world networks are considered as voters. The hierarchy measures are α_c = k-core, α_t = k-truss, α_l = LRC, and α_{tp} = triangle participation. The centrality measures are β_d = Degree, β_l = Local, β_b = Betweenness, β_c = Current-flow Closeness, β_k = Katz, and β_p = PageRank. Note that ties in the output can exist with the Schulze method. In case of ties in the sum of rankings, ranks are broken randomly.

	ρ_s	ρ_p	τ_b	SC_c
1. $\gamma(\alpha_l, \beta_c)$	1	1	2	4/3
2. $\gamma(\alpha_l, \beta_l)$	2	4	4	10/3
3. $\gamma(\alpha_{tp}, \beta_d)$	6	3	3	12/3
4. $\gamma(\alpha_c, \beta_d)$	3	9	1	13/3
5. $\gamma(\alpha_c, \beta_c)$	5	5	5	15/3
6. $\gamma(\alpha_l, \beta_k)$	4	6	7	17/3
7. $\gamma(\alpha_{tp}, \beta_k)$	8	2	8	18/3
8. $\gamma(\alpha_c, \beta_k)$	7	8	6	21/3
9. $\gamma(\alpha_{tp}, \beta_b)$	23	19	17	30/3
10. $\gamma(\alpha_c, \beta_l)$	10	10	10	30/3
11. $\gamma(\alpha_{tp}, \beta_l)$	12	7	12	31/3
12. $\gamma(\alpha_{tp}, \beta_p)$	13	9	10	32/3
13. $\gamma(\alpha_t, \beta_d)$	11	13	8	32/3
14. $\gamma(\alpha_l, \beta_d)$	15	8	11	34/3
15. $\gamma(\alpha_t, \beta_k)$	16	12	10	38/3
16. $\gamma(\alpha_c, \beta_p)$	14	15	10	39/3
17. $\gamma(\alpha_t, \beta_l)$	18	11	13	42/3
18. $\gamma(\alpha_t, \beta_c)$	17	14	13	44/3
19. $\gamma(\alpha_{tp}, \beta_c)$	19	17	9	45/3
20. $\gamma(\alpha_t, \beta_p)$	19	18	14	51/3
21. $\gamma(\alpha_l, \beta_b)$	20	16	15	51/3
22. $\gamma(\alpha_l, \beta_p)$	22	13	18	53/3
23. $\gamma(\alpha_c, \beta_b)$	21	20	16	57/3
24. $\gamma(\alpha_t, \beta_b)$	24	21	19	64/3

samples is provided in table 10. It can be seen that Pearson is the least correlated with values around 0.6 while Spearman and Kendall Tau rankings are well correlated with a value of 0.8. As the correlation between the rankings of the 3 correlation measures is not so low, a final rank is provided based on their average ranking. If we refer to this final rank, the most correlated combinations of centrality and hierarchy are LRC α_l and closeness centrality β_c. It is followed by LRC α_l and local centrality β_l. The third combination is triangle participation α_{tp} and degree centrality β_d. Those combinations have been voted as the most correlated across the 28 real-world networks. For the least correlated combinations, the lowest ranks happen to involve betweenness centrality β_b with the nested hierarchies k-truss α_t and k-core α_c and then comes PageRank β_p with k-truss α_t.

Table 12 reports the Schulze rankings according to the similarity evaluation measures. Kendall Tau correlation between these rankings is provided in table 10. It can be seen that on average the Kendall Tau correlation of the similarity evaluation measures are more correlated than the correlation evaluation measure. Therefore, the rankings based on the five different similarity evaluation measures are averaged in order to obtain a final ranking in table 12. According to this ranking, the top 3 combinations are LRC α_l and closeness centrality β_c. Then comes the combination of triangle participation α_l and Katz centrality β_k followed by LRC α_l and degree centrality β_d. The least similar combinations also involve betweenness centrality β_b with k-core α_c and k-truss α_t, and then comes PageRank β_p with k-core α_c.

TABLE 12. Ranking of the Schulze method For all possible combinations between hierarchy α and Centrality measures β. The ranks are sorted based on the average (SC_s) of all correlation evaluation measures γ used: Jaccard (J), RBO top-k at $p = 0.5$ and $p = 0.9$ ($R^{top}_{0.5}$. $R^{top}_{0.9}$), and RBO of the entire set of nodes at $p = 0.5$ and $p = 0.9$ ($R^{all}_{0.5}$ and $R^{all}_{0.9}$). The 28 real-world networks are considered as voters. The hierarchy measures are α_c = k-core, α_t = k-truss, α_l = LRC, and α_{tp} = triangle participation. The centrality measures are β_d = Degree, β_l = Local, β_b = Betweenness, β_c = Current-flow Closeness, β_k = Katz, and β_p = PageRank. Note that ties in the output can exist with the Schulze method. In case of ties in the sum of rankings, ranks are broken randomly.

	J	$R^{top}_{0.5}$	$R^{top}_{0.9}$	$R^{all}_{0.5}$	$R^{all}_{0.9}$	SC_s
1. $\gamma(\alpha_l, \beta_c)$	1	1	1	1	1	5/5
2. $\gamma(\alpha_{tp}, \beta_k)$	4	3	3	2	2	14/5
3. $\gamma(\alpha_l, \beta_d)$	5	2	2	2	7	18/5
4. $\gamma(\alpha_l, \beta_l)$	2	4	5	4	5	20/5
5. $\gamma(\alpha_{tp}, \beta_l)$	6	6	5	3	4	24/5
6. $\gamma(\alpha_l, \beta_b)$	9	5	4	4	3	25/5
7. $\gamma(\alpha_{tp}, \beta_d)$	3	9	5	4	5	26/5
8. $\gamma(\alpha_l, \beta_k)$	6	7	6	5	6	30/5
9. $\gamma(\alpha_l, \beta_p)$	8	8	6	6	8	36/5
10. $\gamma(\alpha_{tp}, \beta_c)$	7	10	7	7	9	40/5
11. $\gamma(\alpha_{tp}, \beta_p)$	11	11	8	8	10	48/5
12. $\gamma(\alpha_t, \beta_d)$	12	13	11	9	13	58/5
13. $\gamma(\alpha_{tp}, \beta_b)$	16	15	9	10	11	61/5
14. $\gamma(\alpha_t, \beta_k)$	10	12	10	14	17	63/5
15. $\gamma(\alpha_c, \beta_d)$	15	17	12	9	12	65/5
16. $\gamma(\alpha_t, \beta_l)$	13	14	12	13	14	66/5
17. $\gamma(\alpha_c, \beta_k)$	14	16	13	12	16	71/5
18. $\gamma(\alpha_c, \beta_l)$	16	19	17	11	15	78/5
19. $\gamma(\alpha_t, \beta_c)$	16	15	12	17	20	80/5
20. $\gamma(\alpha_c, \beta_c)$	15	20	15	15	18	83/5
21. $\gamma(\alpha_t, \beta_p)$	17	18	14	19	21	89/5
22. $\gamma(\alpha_c, \beta_p)$	18	22	19	16	19	94/5
23. $\gamma(\alpha_t, \beta_b)$	19	21	16	20	23	99/5
24. $\gamma(\alpha_c, \beta_b)$	20	23	18	18	22	101/5

Moreover, in Schulze's similarity rankings, it can be seen in table 12 that flow hierarchy and mixed hierarchy combinations are ranked as top combinations. From rank 1 till rank 11, all combinations involve either LRC flow hierarchy (α_l) or mixed hierarchy triangle participation (α_{tp}). On the other hand, low-ranked combinations involve nested hierarchies k-core (α_c) and k-truss (α_t). This is not the case with correlation where ranks are mixed throughout all different hierarchy types. Nevertheless, comparing the Schulze rankings between both correlation and similarity, there are important common characteristics. Both share the same top combination between LRC flow hierarchy and current-flow closeness centrality. Both don't include any nested hierarchy combination in the top 3. Finally, both share common measures with low ranking (k-core, k-truss, and betweenness centrality). In addition, neither k-core nor k-truss appears in the top 10 combinations in similarity rankings.

In summary, according to the vast majority of rankings, k-core and k-truss are the hierarchy measures which are the less similar/correlated with the centrality measures while LRC and triangle participation appears more frequently as the most similar/correlated with centrality measures. Among the centralities, betweenness and PageRank happen to be the least ones occurring in top rankings as well.

VII. DISCUSSION

The main goal of this study is to disentangle the relationship between hierarchy and centrality measures in complex networks. To do so, multiple evaluation measures have been used in order to evaluate the interplay between various prominent centrality measures and hierarchy measures in a representative sample of real-world networks originating from different fields. In the first set of experiments, for each network S_i, an evaluation measure $\gamma(\alpha_i, \beta_j)$ is calculated for each hierarchy α_i and centrality β_j measure combinations. The evaluation measures include 3 correlation measures (ρ_p, ρ_s, τ_b) and 2 similarity measures (J, RBO). Heatmaps representing the values of the evaluation measures for all the combinations associated with each network showed a wide variability between the hierarchy-centrality combinations on one hand and between the networks on the other hand. Nevertheless, typical behaviors have emerged. First of all, the heatmaps of the correlation and the Jaccard evaluation measures are more patchy as compared to RBO. Additionally, a number of network categories appear where the interactions between centrality and hierarchy measures are quite comparable. One can also notice already that betweenness centrality and k-core with k-truss hierarchy measures stand out from their competitors. At this stage, it appears that there is an underlying organization yet the inner workings between hierarchy, centrality, and network topology need further investigations.

This leads to the second set of experiments in order to further understand the interplay between hierarchy, centrality, and network topology. The goal of these experiments is to cluster the networks according to a given evaluation measure and to investigate if some topological properties are shared by networks belonging to the same cluster. First of all, each network is represented by the sample set of a given evaluation measure, and the Pearson correlation between the sample sets of the networks two-by-two is computed. Hence, if two networks exhibit a similar behavior, the Pearson correlation between their sample sets is high. Results show that some networks share similar behavior. This appears more frequently using the RBO evaluation measure compared to correlation and Jaccard. Indeed, the heatmap representing the Pearson correlation of the networks is less patchy in this case. At this step, more processing is needed in order to cluster the networks. Two strategies are used. First of all, a threshold is used on the correlation matrix in order to distinguish meaningful with non-meaningful correlation values for each evaluation measure. Then, the number of meaningful correlation values is used to rank the networks. Once the networks are ranked, patterns are searched in the topological properties allowing to cluster the networks. Results clearly demonstrate that the driving force explaining the behavior of the various centrality and hierarchy combinations is the density, followed by the transitivity and to a lesser extent assortativity. More specifically, the proportion of meaningful correlation/similarity between hierarchy and centrality combinations is high when both density and transitivity are high. Second, k-means clustering algorithm is used where each network is represented by the sample set of a given evaluation measure. The number of clusters is fixed according to the previous experiment. Overall, it appears that the clusters are quite similar. This consistency strengthens our findings.

The goal of the third set of experiments is to investigate the orthogonality of the various hierarchy and centrality measures under test. Indeed, if centrality and hierarchy measures convey the same information their correlation or similarity is pretty high. The Schulze method allows to rank the various combinations from most correlated/similar to least correlated/similar. Results of these experiments according to the various evaluation measures are quite consistent. LRC and current-flow closeness followed by triangle participation and degree centrality are at the top of the ranking. At the other extreme, k-core and k-truss appear to be the least correlated/similar to betweenness and PageRank centrality. Therefore, the latter should be favored in order to take advantage of the joint information provided by centrality and hierarchy measures.

Now let's relate the values of the evaluation measures to the topological properties of the networks. Considering correlation as an evaluation measure three clusters emerge while there are only two when similarity is used. The top clusters are characterized by a high number of meaningful correlation/similarity between the various combinations of hierarchy and centrality measures. Furthermore, the networks exhibit high density and high transitivity. In this case, centrality and hierarchy exhibit comparable behavior. In other words, high density and transitivity seem to induce that both centrality and hierarchy measures carry the same type of information about the network structure. This is particularly true for LRC and triangle participation that are always well correlated with the centrality measures. In contrast, in the clusters characterized by low density and/or low transitivity k-truss and to a lesser extent k-core behave quite differently than the centrality measures. Indeed, in this situation, very low values of similarity/correlation with the vast majority of centrality measures are commonly observed. The k-truss nested hierarchy is based on the existence of triangles in the network. In case a network has low transitivity and/or density, it is the first measure to account for this feature. Consequently, this results in a greater divergence with centrality measures.

To summarize, k-core and k-truss hierarchy and betweenness and PageRank centrality measures tend to be the least correlated and the least similar as compared to the alternative hierarchies and centrality measures under test. This trend is more pronounced with networks characterized by low density and low transitivity. In networks having high density and high transitivity, it is more difficult to observe a different behavior of hierarchy measures as compared to centrality measures.

VIII. CONCLUSION

The aim of this paper is to identify the relationship between hierarchy and centrality, as both measures are used to quantify the importance of a node. Several experiments have been performed using different correlation and similarity evaluation measures on real-world networks originating from various fields. Results of these investigations reveal precious information about the interactions between centrality measures, hierarchy measures, and macroscopic topological properties of the network. They give a clear guide about which centrality measures and which hierarchy measures should be used in practical applications.

First of all, these investigations demonstrate that centrality and hierarchy measures exhibit a different view of the influence of the nodes.

Second, they show that there is a strong relationship between the hierarchy measures, the centrality measures, and the topological properties of a network. High density and high transitivity tend to bring closer centrality and hierarchy measures in terms of correlation or similarity. Conversely, low values of density and/or transitivity translate into low correlation/similarity between hierarchy and centrality measures.

Third, it appears that the most orthogonal combinations (least similar and least correlated) involve the nested hierarchy measures (k-core and k-truss) and betweenness and PageRank centrality measures. On the opposite, the flow hierarchy and mixed hierarchy measures (LRC and triangle participation) exhibit a comparable behavior than degree centrality and current-flow closeness centrality measures as measured by correlation and similarity measures.

These results of this study have multiple implications. First, it shows that one can substitute a centrality measure to a hierarchy measure in case they are highly correlated and similar. In contrast, combining both measures can also be a good option in order to take into consideration the complementary information carried by each measure when they are not correlated. All of this must be done, taking into consideration the density and transitivity of the given network.

Second, they open new perspectives in the design of effective measures of influence based on these two complementary dimensions. Indeed, they give clear indications about which hierarchy and centrality measures should be combined in a multidimensional scheme in order to design effective influence measures.

Finally, they pave the way for further investigation about the relationship between centrality and hierarchy measures considering complementary information on the network topology such as the community structure.

REFERENCES

[1] A. Baker, "Complexity, networks, and non-uniqueness," *Found. Sci.*, vol. 18, no. 4, pp. 687–705, Nov. 2013.

[2] F. A. Rodrigues, "Network centrality: An introduction," in *A Mathematical Modeling Approach from Nonlinear Dynamics to Complex Systems*. Cham, Switzerland: Springer, 2019, pp. 177–196.

[3] M. Ghanem, C. Magnien, and F. Tarissan, "Centrality metrics in dynamic networks: A comparison study," *IEEE Trans. Netw. Sci. Eng.*, vol. 6, no. 4, pp. 940–951, Oct. 2019.

[4] M. Jalili, A. Salehzadeh-Yazdi, Y. Asgari, S. S. Arab, M. Yaghmaie, A. Ghavamzadeh, and K. Alimoghaddam, "CentiServer: A comprehensive resource, Web-based application and R package for centrality analysis," *PLoS ONE*, vol. 10, no. 11, Nov. 2015, Art. no. e0143111.

[5] S. Das, "On local and global centrality in large scale networks," 2014, *arXiv:1405.5512*. [Online]. Available: http://arxiv.org/abs/1405.5512

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

[7] C. Sciarra, G. Chiarotti, F. Laio, and L. Ridolfi, "A change of perspective in network centrality," *Sci. Rep.*, vol. 8, no. 1, Dec. 2018, Art. no. 15269.

[8] Z. Ghalmane, M. El Hassouni, C. Cherifi, and H. Cherifi, "Centrality in modular networks," *EPJ Data Sci.*, vol. 8, no. 1, p. 15, Dec. 2019.

[9] Z. Ghalmane, C. Cherifi, H. Cherifi, and M. E. Hassouni, "Centrality in complex networks with overlapping community structure," *Sci. Rep.*, vol. 9, no. 1, pp. 1–29, Dec. 2019.

[10] H. Cherifi, G. Palla, B. K. Szymanski, and X. Lu, "On community structure in complex networks: Challenges and opportunities," *Appl. Netw. Sci.*, vol. 4, no. 1, pp. 1–35, Dec. 2019.

[11] H. A. Simon, "The architecture of complexity" *Proc. Amer. Philos. Soc.*, vol. 106, no. 6, pp. 467–482, 1962.

[12] A. Zafeiris and T. Vicsek, *Why We Live in Hierarchies? A Quantitative Treatise*. New York, NY, USA: Springer, 2017.

[13] A. Clark, C. Fox, and S. Lappin, *The Handbook of Computational Linguistics and Natural Language Processing*. Hoboken, NJ, USA: Wiley, 2013.

[14] E. Mones, L. Vicsek, and T. Vicsek, "Hierarchy measure for complex networks," *PLoS ONE*, vol. 7, no. 3, Mar. 2012, Art. no. e33799.

[15] J. Luo and C. L. Magee, "Detecting evolving patterns of self-organizing networks by flow hierarchy measurement," *Complexity*, vol. 16, no. 6, pp. 53–61, Jul. 2011.

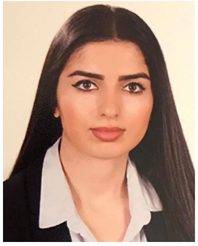
[16] D. Czégel and G. Palla, "Random walk hierarchy measure: What is more hierarchical, a chain, a tree or a star?" *Sci. Rep.*, vol. 5, no. 1, Nov. 2016, Art. no. 17994.

[17] A. Trusina, S. Maslov, P. Minnhagen, and K. Sneppen, "Hierarchy measures in complex networks," *Phys. Rev. Lett.*, vol. 92, no. 17, Apr. 2004, Art. no. 178702.

[18] H. Tangmunarunkit, R. Govindan, S. Jamin, S. Shenker, and W. Willinger, "Network topologies, power laws, and hierarchy," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 32, no. 1, p. 76, 2002.

[19] J. García-Algarra, J. M. Pastor, J. M. Iriondo, and J. Galeano, "Ranking of critical species to preserve the functionality of mutualistic networks using the k-core decomposition," *PeerJ*, vol. 5, p. e3321, May 2017.

[20] J. Wang and J. Cheng, "Truss decomposition in massive networks," 2012, *arXiv:1205.6693*. [Online]. Available: http://arxiv.org/abs/1205.6693
[21] F. D. Malliaros, C. Giatsidis, A. N. Papadopoulos, and M. Vazirgiannis, "The core decomposition of networks: Theory, algorithms and applications," *VLDB J.*, vol. 29, no. 1, pp. 61–92, Jan. 2020.
[22] M. E. J. Newman, *Networks: An Introduction*. New York, NY, USA: Oxford Univ. Press, 2010.
[23] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse, "Identification of influential spreaders in complex networks," *Nature Phys.*, vol. 6, no. 11, pp. 888–893, Nov. 2010.
[24] R. Saxena, S. Kaur, and V. Bhatnagar, "Social centrality using network hierarchy and community structure," *Data Mining Knowl. Discovery*, vol. 32, no. 5, pp. 1421–1443, Sep. 2018.
[25] Y.-Y. Liu, J.-J. Slotine, and A.-L. Barabási, "Control centrality and hierarchical structure in complex networks," *PLoS ONE*, vol. 7, no. 9, Sep. 2012, Art. no. e44459.
[26] Y. Li, W. Li, Y. Tan, F. Liu, Y. Cao, and K. Y. Lee, "Hierarchical decomposition for betweenness centrality measure of complex networks," *Sci. Rep.*, vol. 7, no. 1, Jun. 2017, Art. no. 46491.
[27] C. Li, Q. Li, P. Van Mieghem, H. E. Stanley, and H. Wang, "Correlation between centrality metrics and their application to the opinion model," *Eur. Phys. J. B*, vol. 88, no. 3, pp. 1–13, Mar. 2015.
[28] J. R. F. Ronqui and G. Travieso, "Analyzing complex networks through correlations in centrality measurements," *J. Stat. Mech., Theory Exp.*, vol. 2015, no. 5, May 2015, Art. no. P05030.
[29] D. Schoch, T. W. Valente, and U. Brandes, "Correlations among centrality indices and a class of uniquely ranked graphs," *Social Netw.*, vol. 50, pp. 46–54, Jul. 2017.
[30] S. Oldham, B. Fulcher, L. Parkes, A. Arnatkevičiūtė, C. Suo, and A. Fornito, "Consistency and differences between centrality measures across distinct classes of networks," *PLoS ONE*, vol. 14, no. 7, Jul. 2019, Art. no. e0220061.
[31] D. Lane, "Hierarchy, complexity, society," in *Hierarchy in Natural and Social Sciences*. Dordrecht, The Netherlands: Springer, 2006, ch. 4, pp. 81–119.
[32] T. F. Allen and T. B. Starr, *Hierarchy: Perspectives for Ecological Complexity*. Chicago, IL, USA: Univ. Chicago Press, 2017.
[33] S. Harenberg, G. Bello, L. Gjeltema, S. Ranshous, J. Harlalka, R. Seay, K. Padmanabhan, and N. Samatova, "Community detection in large-scale networks: A survey and empirical evaluation," *Wiley Interdiscipl. Rev., Comput. Statist.*, vol. 6, no. 6, pp. 426–439, Nov. 2014.
[34] J. Yang and J. Leskovec, "Defining and evaluating network communities based on ground-truth," *Knowl. Inf. Syst.*, vol. 42, no. 1, pp. 181–213, Jan. 2015.
[35] J. Mothe, K. Mkhitaryan, and M. Haroutunian, "Community detection: Comparison of state of the art algorithms," in *Proc. Comput. Sci. Inf. Technol. (CSIT)*, Sep. 2017, pp. 125–129.
[36] P. Wagenseller, III, and F. Wang, "Community detection algorithm evaluation using size and Hashtags," 2016, *arXiv:1612.03362*. [Online]. Available: http://arxiv.org/abs/1612.03362
[37] K. Das, S. Samanta, and M. Pal, "Study on centrality measures in social networks: A survey," *Social Netw. Anal. Mining*, vol. 8, no. 1, p. 13, Dec. 2018.
[38] Z. Ghalmane, M. E. Hassouni, and H. Cherifi, "Immunization of networks with non-overlapping community structure," *Social Netw. Anal. Mining*, vol. 9, no. 1, p. 45, Dec. 2019.
[39] A. Ibnoulouafi, M. El Haziti, and H. Cherifi, "M-centrality: Identifying key nodes based on global position and local degree variation," *J. Stat. Mech., Theory Exp.*, vol. 2018, no. 7, Jul. 2018, Art. no. 073407.
[40] N. Gupta, A. Singh, and H. Cherifi, "Centrality measures for networks with community structure," *Phys. A, Stat. Mech. Appl.*, vol. 452, pp. 46–59, Jun. 2016.
[41] Z. Ghalmane, M. E. Hassouni, and H. Cherifi, "Betweenness centrality for networks with non-overlapping community structure," in *Proc. IEEE Workshop Complex. Eng. (COMPENG)*, Oct. 2018, pp. 1–5.
[42] M. Ashtiani, A. Salehzadeh-Yazdi, Z. Razaghi-Moghadam, H. Hennig, O. Wolkenhauer, M. Mirzaie, and M. Jafari, "A systematic survey of centrality measures for protein-protein interaction networks," *BMC Syst. Biol.*, vol. 12, no. 1, p. 80, Dec. 2018.
[43] U. Brandes and D. Fleischer, "Centrality measures based on current flow," in *Proc. Annu. Symp. Theor. Aspects Comput. Sci.* Berlin, Germany: Springer, 2005, pp. 533–544.
[44] W. Webber, A. Moffat, and J. Zobel, "A similarity measure for indefinite rankings," *ACM Trans. Inf. Syst.*, vol. 28, no. 4, pp. 1–38, Nov. 2010.
[45] M. Schulze, "A new monotonic, clone-independent, reversal symmetric, and condorcet-consistent single-winner election method," *Social Choice Welfare*, vol. 36, no. 2, pp. 267–303, Feb. 2011.
[46] A. R. Rossi and K. N. Ahmed, "The network data repository with interactive graph analytics and visualization," in *Proc. AAAI*, 2015, pp. 1–2. [Online]. Available: http://networkrepository.com
[47] V. Latora, V. Nicosia, and G. Russo, *Complex Networks: Principles, Methods and Applications*. Cambridge, U.K.: Cambridge Univ. Press, 2017. [Online]. Available: https://www.complex-networks.net/datasets.html.
[48] W. D. Nooy, A. Mrvar, and V. Batagelj, *Exploratory Social Network Analysis With Pajek: Revised and Expanded Edition for Updated Software*, vol. 46. Cambridge, U.K.: Cambridge Univ. Press, 2018. [Online]. Available: http://vlado.fmf.uni-lj.si/pub/networks/data/esna/default.htm
[49] J. Kunegis, "Handbook of network analysis [KONECT—The Koblenz network collection]," 2014, *arXiv:1402.5500*. [Online]. Available: http://arxiv.org/abs/1402.5500 and http://konect.uni-koblenz.de



STEPHANY RAJEH received the B.Sc. degree in management information systems and the M.Sc. degree in data science for risk analysis from Lebanese University, in 2017 and 2019, respectively. She is currently pursuing the Ph.D. degree with the University of Burgundy, Dijon, France. Her research interests include complex networks, machine learning, and deep learning.

MARINETTE SAVONNET received the Ph.D. degree in computer science, in 1996. She has been an Associate Professor with the University of Burgundy, Dijon, France, since 1996. She is involved in several research projects, which allowed her to establish national and international collaborations often interdisciplinary. Her current research interests include scientific information systems and social networks.

ERIC LECLERCQ received the Ph.D. degree in computer science from the University of Burgundy, in 2000. He is currently an Associate Professor in computer science with the Laboratory of Computer Science (LIB), University of Burgundy. His research interest is in the field of databases that include data models, query languages, and polystore for high performance analytics. He is also involved in complex networks analysis and its applications to social data.

HOCINE CHERIFI (Senior Member, IEEE) received the Ph.D. degree from the National Polytechnic Institute, Grenoble, in 1984. He held faculty positions at Rouen University and Jean Monnet University, France. He has also held visiting positions at Yonsei, South Korea, The University of Western Australia, Australia, National Pintung University, Taiwan, and Galatasaray University, Turkey. He has been a Full Professor of computer science with the University of Burgundy, Dijon, France, since 1999. He has published more than 200 scientific articles in international refereed journals and conference proceedings. His recent research interests include computer vision and complex networks. He held leading positions in more than 15 international conference organizations (General Chair and Program Chair). He served in more than 100 program committees. He is the Founder of the International Conference on Complex Networks and their Applications. He is currently a member of the Editorial Board of *Computational Social Networks*, *PLOS One*, IEEE Access, the *Journal of Imaging*, *Complex Systems*, *Quality and Quantity*, and *Scientific Reports*. He is the Founding Editor-in-Chief of the *Applied Network Science* journal.