IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received October 17, 2020, accepted November 10, 2020, date of publication November 16, 2020, date of current version November 25, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3038306

Cross-Face Centrality: A New Measure for Identifying Key Nodes in Networks Based on Formal Concept Analysis

MOHAMED HAMZA IBRAHIM[1,2], ROKIA MISSAOUI[1], AND JEAN VAILLANCOURT[3]
[1]Département d'informatique et d'ingénierie, Université du Québec en Outaouais, Gatineau, QC J8X 3X7, Canada
[2]Department of Mathematics, Faculty of Science, Zagazig University, Zagazig 44519, Egypt
[3]Department of Decision Sciences, HEC Montreal, Montreal, QC H3T 2A7, Canada

Corresponding author: Rokia Missaoui (rokia.missaoui@uqo.ca)

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

ABSTRACT Discovering influential nodes (or actors) in the network is often the key task of mining, analyzing, and understanding real-life networks. Centrality measures are commonly used to detect important nodes that control the information propagation in the network. While off-the-shelf centrality indices may provide effective node identification in several situations, they frequently produce inadequate results when confronted with massive networks, in the presence of complex local structures or the lack of certain characteristics. In this paper, we introduce *Cross-face*, a new scalable centrality measurement for the identification of key nodes in such networks. Inspired by the Formal Concept Analysis (FCA) framework, the conceptual idea of "Cross-face" is to leverage the faces of concepts to identify nodes that are located in "face bridges" and have an influential "cross clique" connectivity. Thus, it concurrently measures how the node influences its neighbour nodes through its cross cliques while linking the densely connected substructures of the network via its presence in bridges. Unlike traditional centrality measures, the cross-face of nodes can be computed using only a set of symmetrical concepts, which is often quite small compared to the set of nodes or edges in the network. Our experiments on several real-world networks show the efficiency of Cross-face over existing prominent centrality indices such as betweenness, closeness, eigenvector, and k-shell among others.

INDEX TERMS Formal concept analysis, complex networks, key nodes, cross-cliques connectivity, betweenness centrality.

I. INTRODUCTION

A complex system consists of many connected components which non-linearly interact with each other in such a way that the behaviour of its network formulation is often hard to interpret. Since the structure of such network is complex, some nodes are expected to be more important than others in some context. Thus, identifying key nodes in a network is often a substantial task for explaining different behaviours and outcomes in scientific data analysis. This is applicable to a growing number of applications such as halting virus super-spreading, recognizing actionable actors in criminal networks, detecting important infrastructure locations in the Internet networks and predicting active protein modules in protein-protein interaction networks.

The role of nodes within a network can often be deduced through a centrality analysis, which aims to measure how a node influences, or is affected by, other nodes via its connection topology. Since there is no consensus on a unique definition of centrality, many different centrality measures have been proposed in the network analysis literature (cf. [1]–[3] for a detailed survey), each of which considers its specific concept of a central node. That is, some centrality measures may exploit certain valuable characteristics in the network to capture local information, while another measure quantifies how a node is located within a global network context. While there are many classifications of centrality indices based on how their approaches define the importance of a node, a traditional stream research area categorizes the

The associate editor coordinating the review of this manuscript and approving it for publication was Fan-Hsun Tseng.

centrality indices into local and global ones, while taking into account the perception of the information flow of network community properties.

From a cohesion perspective, the local centrality measure focuses on the relative prominence of the focal node in its neighbourhood within local communities. The degree centrality [4] is a simple local measure that counts the number of edges (or links) attached to each node. Thus, the higher the degree of a node, the more important it is in a network. While it is intuitively anticipated that a densely connected node (i.e. a hub) could be in the center, it may not be in practice. For example, the peripheral nodes with the highest degree should not be the central ones in the network. Here, the k-shell centrality [5] provides the k-core, as a recipe measure, to promote the degree of a node in terms of its neighbourhood connections. In k-shell centrality, when removing low degree nodes, the peripheral nodes are not involved in the main connected components, and the high-degree peripheral nodes ought to have a small number of k-cores. This implies that only the hubs at the cores of networks have the highest values of k-cores. The two-fold limitation of the k-shell is that it often assigns the same number of k-cores to many nodes. It also produces poor results when the network structure contains a small number of k-cores, which is often present in real-life networks. In this context, the Cross k-clique connectivity [6], [7] conceptually surpasses k-shell by using k-cliques rather than k-cores. That is, the number of neighbour overlapping k-cliques of a node is counted to determine its connectivity. Although the cross-clique connectivity is a good local measure of how the node facilitates (or influences) the propagation of information through its high overlapping cliques, it has the following two limitations in general:

Limitation a1: Computing the cross-clique connectivity of a node could take an exponential time and space complexity. In its basic form, it depends on the extraction of k-cliques from the network, which is an NP-hard problem. [8].

Limitation a2: The output centrality value of the cross-clique connectivity is often sensitive to the k parameter. That is, a small value of k could lead the cross-clique centrality to overestimate the importance of certain nodes on the periphery of the network, resulting in similar behaviour as the degree centrality. On the contrary, a large value of k might result in the overshooting of several small k-cliques, which in turn, renders the cross-clique connectivity to underestimate the influence of other nodes in dense local communities within the network.

From the network flows perspective, local centrality measures assume that the different types of paths along which information is transferred capture the description of the networks. For instance, the betweenness centrality [9] quantifies the importance of a node by calculating the number of times it is located in the bridge along the shortest paths between the other nodes in the network. Although betweenness may provide a good indicator of how the node perceives the network information flow, it has the following common limitation:

Limitation b1: The calculation of betweenness centrality is computationally expensive. When the network is dense, it requires a time complexity of $O(n^3)$, where n is the number of nodes. Even with the accelerated algorithm proposed by [9] to compute betweenness centrality of a single node, it runs in $O(nm + n^2 \log n)$, where m is the number of edges in the network. This is computationally problematic even with quite medium-sized networks. Using adjacency list representation of the graph, an alternative solution could be to use the depth-first search procedure to extract all bridges in the network, and then approximate the betweenness by counting the number of bridges containing the node. However, this needs at least a time complexity of $O(nm)$, which still represents a computational bottleneck with large networks.

Closeness is another shortest path-based local centrality. It assumes that the important node is an independent one that should be close to other nodes in the network in terms of distance. Intuitively, this is because it is the node that uses the shortest distance to transfer information to all other nodes. Closeness calculates the average distance between the node and all the other nodes, where the distance between the two nodes is the length of the shortest path between them. It is therefore appropriate, in particular, to characterize the spread of information over the shortest distances, but in general, it presents the following limitation.

Limitation b2: Closeness centrality frequently produces poor results on non-spatial networks [10]. Since it depends only on the shortest distances, and due to the small diameter of the networks, the range of variation is too narrow. Given that the largest distance increases with the logarithm of the number of nodes $\log n$, most complex networks may have a small average length of the shortest path. Thus, the ratio between the largest and the minimum distances is $O(\log n)$, presuming that the minimum distance is equal to one. This means that several nodes could have similar values of closeness centrality despite the fact that they could play different roles in the propagation of information. This makes closeness more appropriate for spatial networks whose distance between nodes is high compared to random networks with almost the same number of nodes and edges. Also, and likewise degree, closeness is mostly suitable for connected networks.

On the flip side of the coin, global measures focus on prominence within the context of the network as a whole. For instance, Eigenvector centrality [11] quantifies how a node is important based on its links to other important neighbour nodes. From the viewpoint of walk structures in the network, it approximates the number of traversals of each node through a random walk of infinite length. Thus, when the node is at the center of the network, then it is more accessible than the other nodes. The eigenvector can be seen as a natural extension of degree centrality in a sense that both of them count walks that start and end from the given node. This might support why it has the following limitation:

Limitation b3: Eigenvector centrality produces inaccurate results based on some network structures. When the network structure involves several hubs, most weights of the eigenvector are biased toward a few nodes. In this situation, most of the nodes will have centrality close to zero, which in turn disqualifies the importance of nodes. This is consistent with the explanation proposed in [12]: using a the localize network regime, it was demonstrated that the hub node and its neighbours have the highest eigenvector values, whereas the remaining nodes have the same centrality value equals to $O(1/n)$. Moreover, the Eigenvector centrality may not be able to distinguish the importance of nodes in other regimes where the network has symmetrical dense regions and a small amount of bridges, as it will be explained using a toy example in subsection IV-D.

In this paper, one of our key goals is to bring the three centrality aspects of shortest paths, cohesiveness and impact on important neighbour nodes closer together through the lens of the formal concept analysis framework. That is, to address the limitations (a1-b3) discussed above, we introduce **Cross-face**, a novel approach to centrality for the identification of actionable nodes within complex networks. We achieve this by first reformulating the concept lattice that represents the network in such a way that the so-called symmetrical concepts in the lattice capture the corresponding local structures in the network. We consequently demonstrate that those symmetrical concepts with intent sizes of k greater than 2 can be used to extract k-cliques from the network. Second, we leverage the faces of the symmetrical concepts with intent sizes equal to 2 for detecting new key bridges that we call face-bridges. Unlike traditional cross-clique connectivity centrality, this results in the efficient extraction of key bridges and overlapping k-cliques from the network without being sensitive to the k parameter. Thus, while our experiments here focused on assessing the importance of nodes, the Cross-face approach could easily be applied to other applications, such as finding maximum cliques and bridges in arbitrary graphs and detecting communities.

Furthermore, since the overlapping k-cliques and bridges are powerful sub-graphs commonly used to measure the cohesion and shortest path centrality of nodes in the network, we consequently design our cross-face centrality to identify important nodes that are located in face bridges and cross k-cliques. Dissimilar to betweenness, we consider only the key bridges rather than all bridges. Thus, the overall cross-face quantifies how the node influences its neighbour nodes through its cross-cliques while connecting the densely connected regions of the network through its presence in key bridges. In contrast to closeness and eigenvector, the hybridization of the key bridges and cross-clique connectivity aspects in cross-face centrality enables it to tackle the various topological structures of the network. The computing of the cross-face requires only a set of symmetrical concepts, and is therefore quite fast in practice.

The rest of the paper is organized in the following manner. Section II recalls some basic definitions of FCA and traditional centrality measures. Section III explains our proposed Cross-face centrality for identifying key nodes in further more detail. In Section IV we conduct a thorough experimental study and a discussion. Finally, Section V presents our conclusions and directions for future work.

II. BACKGROUND

This section will briefly review the main concepts that support the comprehension of our proposed centrality measure by using an illustrative example, which is an excerpt of a COVID-19 viral transmission network. As shown in Figure 1, the network is modeled as an undirected graph $\Upsilon = (\mathcal{G}, \mathcal{I})$, where $\mathcal{G}$ is a set of 13 nodes representing individuals, and $\mathcal{I}$ is a set of edges where an edge $(g_i, g_j) \in \mathcal{I}$ links two nodes, $g_i, g_j \in \mathcal{G}$, if they contacted each other.

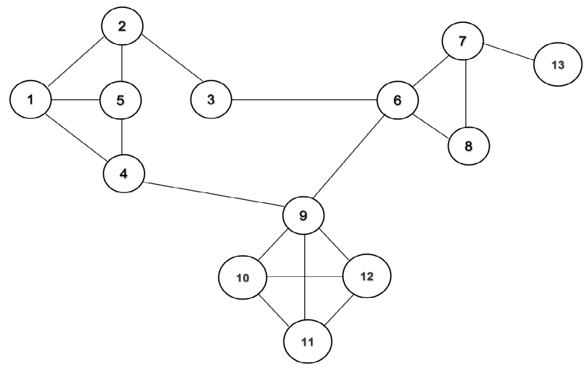

FIGURE 1. A an excerpt of graph network that represents the COVID-19 viral transmission among three local communities.

A. BASIC NOTATION AND DEFINITIONS

Definition 1 (k-Clique): Let $\Upsilon = (\mathcal{G}, \mathcal{I})$ be an undirected graph defined over the objects $\mathcal{G}$. A clique of size k in Υ is a subset $Q \subset \mathcal{G}$ such that $|Q| = k$ and for any two nodes (i.e., objects) g_i and $g_j \in Q$, there exists an edge (i.e., a binary relation) $(g_i, g_j) \in \mathcal{I}$.

For simplicity, we will express a clique by a set of nodes without reference to the edges. For instance, the set $Q = \{9, 10, 11, 12\}$ represents a clique of size 4.

Definition 2 (Bridge or Cut-Edge): An edge $(g_i, g_j) \in \mathcal{I}$ is a bridge *iff* it is not contained in any cycle and its removal increases the number of connected components in the graph Υ.

For example, the edge $\{7, 13\}$ represents a bridge in Υ.

Definition 3 (Local Bridge [13], [14]): An edge $(g_i, g_j) \in \mathcal{I}$ is a local bridge *iff* its removal increases the distance between its two end nodes to a value that is strictly more than two and that there will be no common neighbour nodes between them.

For instance, the edge $\{4, 9\}$ represents a local bridge in Υ. The main difference between local and regular bridges is the number of connected components left behind after their removal. From the network perspective, both regular and local bridges play substantial roles for transmitting the information from one group to another. Thus, we generally use the term "bridge" throughout the paper to denote a local or regular bridge.

B. FORMAL CONCEPT ANALYSIS

In the following we recall key notions of FCA [15] that will be used in this paper.

Definition 4 (Formal Context): It is a triple $\mathbb{K} = (\mathcal{G}, \mathcal{M}, \mathcal{I})$, where $\mathcal{G}$ is a set of objects, $\mathcal{M}$ a set of attributes, and $\mathcal{I}$ a binary relation between $\mathcal{G}$ and $\mathcal{M}$ with $\mathcal{I} \subseteq \mathcal{G} \times \mathcal{M}$. For $g \in \mathcal{G}$ and $m \in \mathcal{M}$, $(g, m) \in \mathcal{I}$ holds (i.e., $(g, m) = 1$) iff the object g has the attribute m, and otherwise $(g, m) \notin \mathcal{I}$ (i.e., $(g, m) = 0$).

Given arbitrary subsets $A \subseteq \mathcal{G}$ and $B \subseteq \mathcal{M}$, the following derivation operators are defined:

$$A' = \{m \in \mathcal{M} \mid \forall g \in A, (g, m) \in \mathcal{I}\}, \quad A \subseteq \mathcal{G}$$
$$B' = \{g \in \mathcal{G} \mid \forall m \in B, (g, m) \in \mathcal{I}\}, \quad B \subseteq \mathcal{M}$$

where A' is the set of attributes common to all objects of A and B' is the set of objects sharing all attributes from B. The closure operator $(.)''$ implies the double application of $(.)'$, which is extensive, idempotent and monotone. The subsets A and B are closed when $A = A''$, and $B = B''$.

Definition 5 (Formal Concept): The pair $c = (A, B)$ is called a *formal concept* of $\mathbb{K}$ with *extent* A and *intent* B if both A and B are closed and $A' = B$, and $B' = A$.

Definition 6 (Partial Order Relation $\preceq$): A concept $c_1 = (A_1, B_1) \preceq c_2 = (A_2, B_2)$ if:

$$A_1 \subseteq A_2 \iff B_1 \supseteq B_2. \tag{1}$$

In this case, c_2 is called a *superconcept* (or successor) of c_1, and c_1 is called a *subconcept* (or predecessor) of c_2. The set of all concepts of the formal context $\mathbb{K}$ is expressed by $\mathcal{C}(\mathbb{K})$ or simply $\mathcal{C}$.

Definition 7 (Concept Lattice): The concept lattice of a formal context $\mathbb{K}$, denoted by $\mathcal{L}(\mathbb{K}) = (\mathcal{C}, \preceq)$, is a Hasse diagram that represents all formal concepts $\mathcal{C}$ together with the partial order that holds between them. In $\mathcal{L}(\mathbb{K})$, each node represents a concept with its extent and its intent while the edges represent the partial order between concepts.

There are several methods (cf. [15]–[17]) that build the lattice, i.e., compute all the concepts together with the partial order. One-mode data networks contain only one type of nodes and relations. Hence, we can simply adapt the formal context (in Definition 4) to define a one-mode data context as follows.

Definition 8 (One-Mode Formal Context): It is a formal context $\tilde{\mathbb{K}} = (\mathcal{G}, \mathcal{G}, \mathcal{I})$ in which the two sets of objects and attributes are identical, i.e., $\mathcal{G} \equiv \mathcal{M}$, and $\mathcal{I}$ is a set of relations defined on $\mathcal{G}$ with $\mathcal{I} \subseteq \mathcal{G} \times \mathcal{G}$. For $g_i, g_j \in \mathcal{G}$, $(g_i, g_j) \in \mathcal{I}$ holds iff object g_i is linked to g_j or $g_i = g_j$.

Definition 9 (Lower- and Upper-Covers): For any two formal concepts $c_1 = (A_1, B_1) \preceq c_2 = (A_2, B_2)$ if:

$$(A_1, B_1) \preceq (A_2, B_2), \nexists\, c3 = (A_3, B_3) \text{ such that}$$
$$(A_1, B_1) \preceq (A_3, B_3) \preceq (A_2, B_2), \tag{2}$$

or

$$A_1 \subseteq A_3 \subseteq A_2 \iff B_1 \supseteq B_3 \supseteq B_2, \tag{3}$$

then $c_1 = (A_1, B_1)$ is a *lower cover* of $c_2 = (A_2, B_2)$, and $c_2 = (A_2, B_2)$ is an *upper cover* of $c_1 = (A_1, B_1)$; represented as $c_1 \prec c_2$ and $c_2 \succ c_1$ respectively.

We will use $\mathcal{U}(c)$ to denote the set of upper covers of the formal concept c.

Definition 10 (Concept Face [18]): The *face* $f(c, c_u)$ of a concept $c = (A, B)$ w.r.t its u-th upper cover concept, $c_u = (A_u, B_u) \in \mathcal{U}(c)$, is the difference between their intent sets as: $f(c, c_u) = B \setminus B_u$.

C. CENTRALITY MEASURES

Definition 11 (Centrality Measure): The *centrality measure* of a regular node $g_i \in \mathcal{G}$ is a function that assigns a positive real number to g_i quantifying its centrality w.r.t to all other nodes $\mathcal{G} \setminus \{g_i\}$ in the network Υ.

The centrality measures are frequently used to identify and rank key nodes in networks. While several centrality measures have been introduced, the degree, closeness, betweenness and eigenvector have been found to be the most prominent in several applications, and they thereby are commonly used.

Definition 12 (Degree Centrality D_c [4]): The degree centrality of a node g_i, in a graph network Υ, is defined as:

$$\mathrm{D_c}(g_i) = \sum_{g_j \in \mathcal{G}} I_{ij}, \tag{4}$$

where I_{ij} is equal to 1 when a link exists between g_i and g_j, and 0 otherwise. Thus, the summation in Eq. (4) represents the number of edges (or ties with neighbour nodes) involving the node g_i.

Definition 13 (Closeness Centrality C_c [19]): The closeness centrality of a node g_i, in a graph network Υ, is defined as:

$$\mathrm{C_c}(g_i) = \frac{1}{\sum_{g_j \in \mathcal{G}} d(g_i, g_j)}, \tag{5}$$

where $d(g_i, g_j)$ is the length of (or the number of hops on) the shortest path between the nodes g_i and g_j.

Definition 14 (Betweenness Centrality B_c [9]): The betweenness centrality of a node g_i, in a graph network Υ, is defined as:

$$\mathrm{B_c}(g_i) = \sum_{g_j \neq g_k \neq g_i, \quad g_j, g_k, g_i \in \mathcal{G}} \frac{\sigma_{g_j g_k}(g_i)}{\sigma_{g_j g_k}}, \tag{6}$$

where $\sigma_{g_j g_k}$ denotes the total number of shortest paths between nodes g_j and g_k, and $\sigma_{g_j g_k}(g_i)$ is the number of those paths that traverse g_i.

Definition 15 (Eigenvector Centrality EV_c [11]): The eigenvector centrality of a node g_i, in a graph network Υ, can be iteratively computed as:

$$\mathrm{EV_c}(g_i) = \frac{1}{\lambda} \sum_{g_j \in \mathcal{G}} a_{g_i g_j} \mathrm{EV_c}(g_j), \tag{7}$$

where the eigenvalue $\lambda \neq 0$ is a constant, and $a_{g_i g_j}$ is the adjacency element which is equal to 1 if node g_i is linked to node g_j, and 0 otherwise.

III. CROSS-FACE APPROACH BASED ON FCA

At a conceptual level, our overall Cross-face centrality approach contains the following key elements. First, we build the formal context and construct the concept lattice of the network. Second, we extract from the lattice the set of concepts that represent cross k-cliques and face bridges in the social graph. Third, we calculate the Cross-face centrality measure to identify key nodes.

A. CONSTRUCTING THE FORMAL CONTEXT OF A NETWORK

We first build the formal context of the social network $\Upsilon = (\mathcal{G}, \mathcal{I})$ by computing the symmetrical *modified adjacency matrix* [20] as follows:

$$\tilde{\mathbb{K}} = (\mathcal{G}, \mathcal{G}, \mathcal{I}) = \begin{cases} (g_i, g_j) = 1 & \text{If } \exists\, (g_i, g_j) \in \mathcal{I},\ i \neq j \\ (g_i, g_j) = 1 & \text{if } i = j \\ (g_i, g_j) = 0 & \text{Otherwise.} \end{cases} \tag{8}$$

In Eq. (8), we assign 0 to the element of $\tilde{\mathbb{K}}$ in the row i and column j if the object (node) g_i is not linked to the object g_j in the network Υ. Otherwise, we assign 1 to it. For example, the constructed formal context $\tilde{\mathbb{K}}$ of our toy graph in Figure 1 is represented in Table 1. We then construct the concept lattice $\tilde{\mathbb{K}}$ from the formal context, as it is shown in Figure 2 (*left*).

TABLE 1. The formal context $\tilde{\mathbb{K}}$ for the toy network of Figure 1.

$\mathcal{G}$	1	2	3	4	5	6	7	8	9	10	11	12	13
1	1	1	0	1	1	0	0	0	0	0	0	0	0
2	1	1	1	0	1	0	0	0	0	0	0	0	0
3	0	1	1	0	0	1	0	0	0	0	0	0	0
4	1	0	0	1	1	0	0	0	1	0	0	0	0
5	1	1	0	1	1	0	0	0	0	0	0	0	0
6	0	0	1	0	0	1	1	1	1	0	0	0	0
7	0	0	0	0	0	1	1	1	0	0	0	0	1
8	0	0	0	0	0	1	1	1	0	0	0	0	0
9	0	0	0	1	0	1	0	0	1	1	1	1	0
10	0	0	0	0	0	0	0	0	1	1	1	1	0
11	0	0	0	0	0	0	0	0	1	1	1	1	0
12	0	0	0	0	0	0	0	0	1	1	1	1	0
13	0	0	0	0	0	0	1	0	0	0	0	0	1

B. EXTRACTING CROSS CLIQUES AND FACE BRIDGES

By analyzing the constructed lattice $\mathcal{L}(\tilde{\mathbb{K}})$, it is possible to spot certain concepts within which the intent is equal to the extent, which can be defined as:

Definition 16 (Symmetrical Concept): A formal concept $c = (A, B)$, with extent A and intent B, is called a *symmetrical concept* if $A = B$, i.e., its extent is identical to its intent.

We use $\tilde{C}$ to denote the set of all the *symmetrical concepts* in a network Υ.

Proposition 1: Given a network Υ and its corresponding concept lattice $\mathcal{L}(\tilde{\mathbb{K}})$, a symmetrical concept $\tilde{c} = (A_{\tilde{c}}, B_{\tilde{c}}) \in \mathcal{L}$ with $A_{\tilde{c}} \equiv B_{\tilde{c}}$ and $|A_{\tilde{c}}| = k > 2$, represents a k-clique $Q = \{g_i : g_i \in A_{\tilde{c}}\}$ in Υ.

Proof: A symmetrical concept represents a unit square matrix of size k - as a sub-matrix of the modified adjacency matrix - and hence a k-clique since it is a maximal square in the formal context. Assume now that $Q = \{g_i\}_{i=1}^{k}$ is a k-clique of Υ with $k > 2$. Then, from Definition 1, for any two objects $g_i, g_j \in Q$, there exists an edge (g_i, g_j) in Υ that links the two objects. Based on Eq. (8), the obtained $k \times k$ modified adjacency matrix $\tilde{\mathbb{K}}(Q, Q, \mathcal{I}_Q)$ that expresses the clique Q obviously represents a matrix consisting of all 1's. Such a matrix coincides with the symmetrical concept $\tilde{c} = (\{g_i\}_{i=1}^{k}, \{g_i\}_{i=1}^{k})$ in which both extent $A_{\tilde{c}}$ and intent $B_{\tilde{c}}$ involve only the objects $\{g_i\}_{i=1}^{k}$ of Q. This entails that a k-clique that contains the node set $Q = \{g_i : g_i \in A\}$ is identical to a symmetrical concept $\tilde{c} = (A_{\tilde{c}}, B_{\tilde{c}})$ such that $A_{\tilde{c}} \equiv B_{\tilde{c}} = \{g_i\}_{i=1}^{k}$. ■

Proposition 2: Given a network Υ and its corresponding concept lattice $\mathcal{L}(\tilde{\mathbb{K}})$, a symmetrical concept $\tilde{c} = (A_{\tilde{c}}, B_{\tilde{c}}) \in \mathcal{L}$, with $A_{\tilde{c}} \equiv B_{\tilde{c}} = \{g_i, g_j\},\ |A| = 2$ has at most two upper covers and represents a corresponding bridge (g_i, g_j) in Υ.

Proof: The proposition is held once we prove that: (1) a symmetrical concept $\tilde{c} = (A_{\tilde{c}}, B_{\tilde{c}})$, with $A_{\tilde{c}} \equiv B_{\tilde{c}} = \{g_i, g_j\},\ |A| = 2$ is a bridge if it has at most two upper covers, (2) a bridge is represented by a symmetrical concept with an extent and intent involving only the two objects of the bridge.

For the first part, suppose that $c_u = (A_{c_u}, B_{c_u})$ is an upper cover of $\tilde{c}$. From Definition 9, we have $B_{c_u} \subseteq B_{\tilde{c}} = \{g_i, g_j\}$. This means that B_{c_u} can be equal to one of the four possibilities $\emptyset$, $\{g_i\}$, $\{g_j\}$ or $\{g_i, g_i\}$. If $B_{c_u} = \emptyset$, then c_u is the Supremum. This actually implies that the nodes $\{g_i, g_i\} \in B_{\tilde{c}}$ are not linked to any other nodes in the graph, which in turn means that $\tilde{c}$ is not a bridge. Thus, this possibility, i.e., $B_{c_u} = \emptyset$, contradicts with the assumption that $\tilde{c}$ represents a bridge. If $B_{c_u} = \{g_i, g_i\}$, then we have that $B_{c_u} = B_{\tilde{c}}$ and $A_{c_u} = A_{\tilde{c}}$. This means that $c_u \equiv c_u$, which can not also occur since it contradicts with the lattice theory, i.e., it is not possible that a concept and its upper cover have symmetrical extents and intents. Now, we still have only two possibilities $\{g_i\}$ and $\{g_j\}$ that can represent upper covers of $\tilde{c}$. This implies that $\tilde{c}$ is a bridge when it has at most two upper covers, each of which has an intent that is equal to either $\{g_i\}$ or $\{g_j\}$.

For the second part, let $B_{\tilde{c}} = (g_i, g_j)$ be a bridge between two components $\mathcal{T}_i$ and $\mathcal{T}_j$ of Υ such that $g_i \in \mathcal{T}_i$ and $g_j \in \mathcal{T}_j$. From Eq. (8), the 2×2 modified adjacency matrix of $B_{\tilde{c}}$ defines a unit matrix $J_{B_{\tilde{c}}}$. Now, since each object of the bridge $B_{\tilde{c}}$ belongs to a different component, then its $J_{B_{\tilde{c}}}$ matrix represents also a sub-matrix in the one-mode formal context $\tilde{\mathbb{K}}$ such that: (i) $(g_i, g_p) = 0\ \forall g_i \in \mathcal{T}_i$ and $g_p \in \mathcal{T}_j \backslash \{g_j\}$ and (ii) $(g_j, g_p) = 0\ \forall g_j \in \mathcal{T}_j$, and $g_p \in \mathcal{T}_i \setminus \{g_i\}$. This indeed implies that the modified adjacency matrix $J_{B_{\tilde{c}}}$ of the bridge can be used to extract from $\tilde{\mathbb{K}}$ a symmetrical concept $\tilde{c} = (\{g_i, g_j\}, \{g_i, g_j\})$ where both its intent and extent contain the two nodes of the bridge. ■

Now before explaining what is a face bridge, the first thing we do is to formulate what are non-influential (or non-actionable) nodes from the viewpoint of FCA.

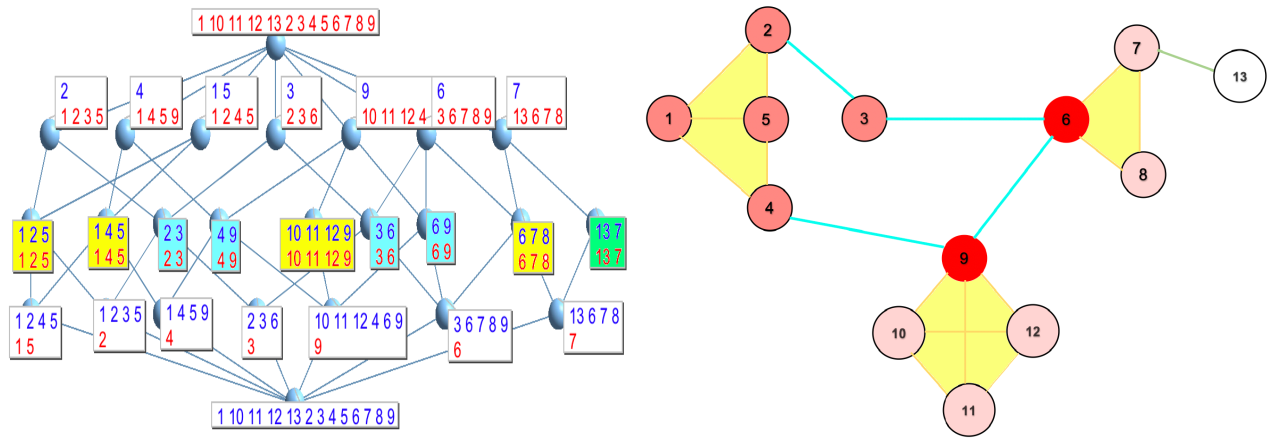

FIGURE 2. ***(left)*** **The concept lattice $\mathcal{L}(\tilde{\mathbb{K}})$ of our example network in Figure 1. The symmetrical concepts which appear in yellow, cyan and green capture the k-cliques, key bridges and non-influential bridges respectively.** ***(right)*** **Their corresponding k-cliques and face-based bridges in the graph network of Figure 1. Pursuant to the Cross-face centrality, the most influential nodes for transmitting the virus are {6, 9}, the next less important ones are {1, 2, 3, 4, 5}, the second less important ones are {7, 8, 10, 11, 12} and the least important node is {13} which has a cross-face centrality equals to 0. The darkness degree of red color reflects the influential rank of the node.**

Definition 17 (Non-Influential Node): For a symmetrical formal concept $\tilde{c}_i = (A_{\tilde{c}_i}, B_{\tilde{c}_i}) \in \tilde{\mathcal{C}}$, a node $g \in B_{\tilde{c}_i}$ is *non-influential* if its removal from $\tilde{c}_i$ (and accordingly from the graph $\mathcal{G}$) does not violate the closure conditions of other symmetrical concepts $\tilde{\mathcal{C}} \setminus \{\tilde{c}_i\}$ that involve it:

$$\forall \tilde{c}_j \in \tilde{\mathcal{C}} \setminus \{\tilde{c}_i\} \text{ and } g \in B_{\tilde{c}_j}, (B_{\tilde{c}_j} \setminus \{g\})'' = B_{\tilde{c}_j}. \tag{9}$$

That is, the subset of symmetrical concepts that contain node g still maintain their local conceptual structures even after removing g from their extents and intents. Intuitively, this means that the node g is not important since taking it off from the graph $\mathcal{G}$ does not affect the essential connectivity of the network (e.g., the collapsing of other symmetrical concepts). For example, in our toy graph of Figure 1, the node 13 is non-influential in the symmetrical concept $\tilde{c} = \big(\{7, 13\}, \{7, 13\}\big)$. This is because its removal from $\mathcal{G}$ does not result in the loss of connectivity of other nodes in the graph.

In fact, the definition 17 raises an interesting question of how to determine the non-influential nodes in a given symmetrical concept? Fortunately, the faces of the symmetrical concept, w.r.t. its the upper covers, can provide it with pieces of information as to what its non-influential nodes would be. For example, by contrasting the intent of the symmetrical concept $\tilde{c} = (\{2, 3\}, \{2, 3\})$ with the intent of its upper cover $\tilde{c}_u = (\{1, 2, 3, 5\}, \{2\})$ we can infer that the object $\{2\}$ is key object in $\tilde{c}$. Thus, an effective strategy here to answering this question is to contrast the symmetrical concept with its upper covers through faces to identify its potential non-influential nodes. That is, the set of faces of a symmetrical concept $\tilde{c} = (A_{\tilde{c}}, B_{\tilde{c}})$, w.r.t. its upper covers $\mathcal{U}(\tilde{c})$, share the non-influential nodes in its intent $B_{\tilde{c}}$:

$$\forall g \in \{\cap_{c_u \in \mathcal{U}(\tilde{c})} f(\tilde{c}, c_u)\} \implies (B_{\tilde{c}_j} \setminus \{g\})'' = B_{\tilde{c}_j}, \\ \forall \tilde{c}_j \in \tilde{\mathcal{C}} \setminus \{\tilde{c}\} \text{ and } g \in B_{\tilde{c}_j}. \tag{10}$$

For instance, the symmetrical concept $\big(\{3, 6\}, \{3, 6\}\big)$ has two faces $f_1(\{3, 6\}, \{6\}) = \{3\}$ and $f_2(\{3, 6\}, \{3\}) = \{6\}$ respectively. The intersection of these two faces is empty, which means that there is no non-influential nodes in the symmetrical concept $\big(\{3, 6\}, \{3, 6\}\big)$. On the contrary, the symmetrical concept $\big(\{7, 13\}, \{7, 13\}\big)$ has only one face $f_1(\{7, 13\}, \{7\}) = \{13\}$, indicating that it has a non-influential node 13.

On the basis of proposition 2 and Equation 10, we can leverage the faces of symmetrical concepts, with intent size 2, to formulate the key bridge[1] by defining it as follows:

Definition 18 (Face Bridge): Given a social graph Υ and its corresponding concept lattice $\mathcal{L}(\mathbb{K})$, a symmetrical concept $\tilde{c} = (A_{\tilde{c}}, B_{\tilde{c}}) \in \mathcal{L}$, with $A_{\tilde{c}} \equiv B_{\tilde{c}} = \{g_i, g_j\}$, $|A_{\tilde{c}}| = 2$, is called a *face bridge* $\{g_i, g_j\}$ in Υ iff there is no intersection among all its faces as:

$$\cap_{d \in \mathcal{U}(\tilde{c})} f(\tilde{c}, c_d) = \emptyset. \tag{11}$$

In Eq. (11), the intersection of the faces informs the symmetrical concept $\tilde{c}$ what are its non-influential objects. If the intersection results in an empty set, then the two end nodes of its corresponding bridge are key. As such, it is a key bridge, and we thereby denote it as a face bridge. Note that, based on Proposition 2, a symmetrical concept, with intent size 2, must have at most two upper covers, i.e., $|\mathcal{U}(\tilde{c})| \leq 2$, and in turn possesses at most two faces. Now, since each face must involve at least one node, then a key property is that the condition in Eq. (11) can not be satisfied when $\tilde{c}$ is a meet-irreducible concept. This because the meet-irreducible concept often has only one face. Thus, we could leverage the meet-irreducible property to accelerate the process of identifying face bridges. That is, instead of computing the faces of the symmetrical concept as in Eq. (11), we can simply

[1]Note that the bridge is key when none of its two end nodes is non-influential.

verify whether it is not a meet-irreducible one that has two upper covers (i.e., two faces) or not:

$$\mathbb{1}_{\tilde{c}} = \begin{cases} 1, & |\mathcal{U}(\tilde{c})| = 2 \\ 0, & \text{otherwise.} \end{cases} \tag{12}$$

where, in Eq. (12), we specify the extracted symmetrical concept with intent size 2 as a face bridge by validating whether it is not a meet-irreducible, i.e., it contains two faces. That is, a symmetrical concept with intent size 2 that has two faces is a face bridge by default. On the contrary, it is certainly not a face bridge when it is meet-irreducible, i.e., has only one face.

In the sequel, we use $\tilde{\mathcal{C}}_{>2}$ and $\tilde{\mathcal{C}}_{=2}$ to denote the subsets of symmetrical concepts that represent k-cliques and face bridges, respectively. For instance, as shown in the lattice of Figure 2 (*left*), $\tilde{\mathcal{C}}_{>2}$ contains four symmetrical concepts $\big(\{1, 2, 5\}, \{1, 2, 5\}\big)$,$\big(\{1, 4, 5\}, \{1, 4, 5\}\big)$,$\big(\{6, 7, 8\}, \{6, 7, 8\}\big)$, and $\big(\{9, 10, 11, 12\}, \{9, 10, 11, 12\}\big)$ that represent four k-cliques $\{1, 2, 5\}, \{1, 4, 5\}, \{6, 7, 8\}, \{9, 10, 11, 12\}$, while $\tilde{\mathcal{C}}_{=2}$ includes four symmetrical concepts $\big(\{2, 3\}, \{2, 3\}\big)$, $\big(\{4, 9\}, \{4, 9\}\big)$, $\big(\{3, 6\}, \{3, 6\}\big)$, and $\big(\{6, 9\}, \{6, 9\}\big)$ that represent four face bridges in the graph network of Figure 1. Note that the bridge $\{7, 13\}$ is not a face bridge since its corresponding symmetrical concept $\big(\{7, 13\}, \{7, 13\}\big)$ violates the indicator function of Eq. (12), by having only one upper cover. At this point, we have paved the way for Cross-face centrality.

C. CROSS-FACE CENTRALITY

Definition 19 (Cross-Face Centrality CF_c): The Cross-face centrality of a node (or object) $g \in \mathcal{G}$, in a given graph network Υ, can be computed as:

$$\mathrm{CF_c}(g) = \overbrace{\frac{|\{\tilde{c} \in \tilde{\mathcal{C}}_{>2} \mid g \in A_{\tilde{c}}\}|}{|\tilde{\mathcal{C}}_{>2}|}}^{\text{Cross-cliques containing } g} + \overbrace{\frac{|\{\tilde{c} \in \tilde{\mathcal{C}}_{=2} \mid g \in A_{\tilde{c}},\ \mathbb{1}_{\tilde{c}} = 1\}|}{|\tilde{\mathcal{C}}_{=2}|}}^{\text{Face-bridges containing } g}. \tag{13}$$

That is, in Eq. 13, the Cross-face centrality computes the sum of *Cross-clique*[2] and *Face-bridge* terms. The numerator of the cross-clique one simply counts the number of symmetrical concepts, with intent sizes greater than 2, that involve the node g. Thus, it quantifies the portion of cross k-cliques, in the graph network Υ, which the node g belongs to. From a conceptual perspective, the cross-clique term can be considered as an efficient way of computing the cross-clique connectivity [6], [7] of the node g. In the face-bridge term, we count the number of the face bridges that involve the node g. Thus, it quantifies the portion of the key bridges that contain the node g in the graph. Note that the numerators of both Cross-clique and Face-bridge terms are unnormalized quantities. Thus, the denominators in Eq. 13 serve as normalization constants to scale the two terms between 0 and 1.

[2]Note that the cross-clique of a node is the number of overlapping cliques to which it belongs.

At a high level, the intuition behind the Cross-face centrality of a node is two-fold. First, it measures how the node locally influences its neighbour nodes through its existence in the cross cliques. Second, it quantifies whether the node is globally crucial in linking densely connected regions of the network through its presence in key bridges. Figure 2 (*right*) makes it easier to understand the guiding idea of the cross-face centrality. It initially nominates the nodes $\{2, 4, 6, 9\}$ as the highest influential ones. This is because those nodes are involved in both cross cliques and face bridges. It then picks out from the list $\{2, 4, 6, 9\}$ the nodes that exist in a higher number of both cross-cliques and face bridges, and it thereby specifies $\{6, 9\}$ as the most influential ones. Other nodes such as $\{1, 3, 5\}$ and $\{7, 8, 10, 11, 12\}$ are designated in the next less-levels of importance since they are evenly involved in either cross-cliques or face bridges. The node $\{13\}$ is involved in neither cross-cliques nor face bridges, and the cross-face thereby does not suggest it as a key node.

Algorithm 1 gives the pseudo-code for computing the Cross-face centrality of all nodes in the graph Υ. The algorithm takes as input the set of all extracted symmetrical concepts $\tilde{\mathcal{C}}$. For each node $g_i \in \mathcal{G}$, it iteratively counts the number of cross k-cliques in $\tilde{\mathcal{C}}$ that involve g_i (lines 5-11). It then counts the number of face bridges that involve g_i (lines 12-18). Subsequently, it computes the Cross-face centrality $\mathrm{CF_c}(g_i)$ of a node g_i (line 19). Finally, it returns a list containing the Cross-face centrality measures $\mathrm{CF_c}$ of all nodes in the graph (line 21). Note that, based on the obtained Cross-face centrality list, we can rank the nodes in a descending order according to their importance. For instance, as shown in Figure 2 (*right*), based on the list of cross-face centrality for our graph of Figure 1, the most key nodes for transmitting the virus are $\{6, 9\}$, the next less important ones are $\{1, 2, 3, 4, 5\}$, the second less important ones are $\{7, 8, 10, 11, 12\}$ and the least important node is $\{13\}$ which has a cross-face centrality equals to 0.

Computational Complexity: In algorithm 1, we need to store all nodes and symmetrical concepts. Thus, it requires a space complexity of $O(n_{\tilde{c}} + n)$, where $n_{\tilde{c}} = |\tilde{\mathcal{C}}|$ is the number of symmetrical concepts.[3] Then, for each node in $\mathcal{G}$, we iterate through all symmetrical concepts in $\tilde{\mathcal{C}}$ to count the number of cross-cliques and face bridges that contain the node. Thus, it takes a time complexity of $O(n \times n_{\tilde{c}})$ for computing the Cross-face centrality of all nodes in a given graph network.

IV. EXPERIMENTAL EVALUATION

The goal of our experimental evaluation was to investigate the following key questions.

- (**Q1**) Is the Cross-face centrality more accurate than state-of-the-art centrality measures?
- (**Q2**) Is cross-face centrality performing fast compared to prominent centrality measures?

[3]Note that $n_{\tilde{c}}$ is equal, at most, to $O(n)$ in the worst case scenario of the network.

Algorithm 1 Computing Cross-Face Centrality (CF_c) for All Nodes in Network

Input: Set of symmetrical concepts ($\tilde{\mathcal{C}}$).
Output: Cross-face centrality measures (CF_c) of all nodes.

$CF_c \leftarrow \emptyset$
for $i \leftarrow 1$ to $|\mathcal{G}|$ **do**
  $\text{count}_1 \leftarrow 0;\ \text{count}_2 \leftarrow 0$
  $n_{\tilde{c}=2} \leftarrow 0;\ n_{\tilde{c}>2} \leftarrow 0$
  for each concept $\tilde{c}_j = (A_j, B_j) \in \tilde{\mathcal{C}}$ **do**
    // Counting cross-cliques that contain the node g_i
    if $|A_j| > 2$ **then**
      $n_{\tilde{c}>2} \leftarrow n_{\tilde{c}>2} + 1$
      if $g_i \in A_j$ **then**
        $\text{count}_1 \leftarrow \text{count}_1 + 1$
      end if
    end if
    // Counting face bridges that contain the node g_i
    if $|A_j| == 2$ and $\mathbb{1}_{\tilde{c}} == 1$ **then**
      $n_{\tilde{c}=2} \leftarrow n_{\tilde{c}=2} + 1$
      if $g_i \in A_j$ **then**
        $\text{count}_2 \leftarrow \text{count}_2 + 1$
      end if
    end if
  end for
  $CF_c[i] \leftarrow (\text{count}_1 / n_{\tilde{c}>2}) + (\text{count}_2 / n_{\tilde{c}=2})$
end for
Return CF_c

- (**Q3**) Is the Cross-face centrality approach correlated to other state-of-the-art centrality measures?

To find robust answers, we first selected the following four real-world social networks which have different complex structures, and they thereby facilitate the validation of various scenarios.

A. DATASETS

- **Email** [21], which is a network of e-mail interchanges between members of the University Rovira i Virgili (Tarragona). Each node represents a user and an edge indicates that two users have exchanged emails.
- **Netscience** [22], which is a co-authorship network of scientists working on network theory and experiment.
- **USAir97** [23], which is a North American transportation network. The nodes represent airports and the edges represent routes between airports.
- **Jazz musicians** [24], which is a collaboration network between Jazz musicians. Each node represents a jazz musician and each edge denotes a cooperation between two musicians in a band.

A brief statistics of the networks is summarized in Table 2.[4]

[4]Datasets are available at: http://www-personal.umich.edu/ mejn/netdata/ https://github.com/gephi/gephi/wiki/Datasets

TABLE 2. A brief statistics of the social networks, which includes the number of nodes ($|\mathcal{G}|$), the number of edges ($|\mathcal{I}|$), the average degree (Φ), the average clustering coefficient (Ψ), the average shortest path length (Ω) and the density (Θ).

Name	$\|\mathcal{G}\|$	$\|\mathcal{I}\|$	Φ	Ψ	Ω	Θ
Email	1133	5451	9.62	0.22	3.61	0.09
Netscience	1589	2742	3.45	0.88	5.82	0.02
USAir97	332	2126	12.80	0.75	2.74	0.38
Jazz musicians	198	2742	27.70	0.63	2.21	0.141

B. METHODOLOGY

Subsequently, we compared the results of our proposed **Cross-face** centrality with the following measures:

- **Closeness** [19], a prominent diameter-based centrality
- **Betweenness** [9], a state-of-the-art geodesics-based centrality
- **Eigenvector** [11], a state-of-the-art centrality that assesses the importance of a node based on its connections to other highly influential nodes in a network.
- **k-shell** [5], a state-of-the-art decomposition-based centrality measure that evaluates the importance of node according to its location within the network. In k-shell, the decomposition process is repeatedly used to remove all nodes with a degree of less than k, where k is incrementally increased from the value 1. As a result, the k-shell value is assigned to the inner nodes on the k-th layer after discarding the outer nodes. The process of decomposition is completed when all nodes are removed. At a high level, the k-shell provides the coarse-grained importance of nodes such that the inner ones often have a high influence. The detailed explanation for k-shell formula has been elegantly provided in [25], [26].
- **Heatmap** [27], a recently proposed centrality measure that efficiently identifies influential nodes in the network by comparing the farness (i.e., the global network information) of a node with the average sum of farness of its adjacent nodes (i.e., the local network information). The Heatmap centrality of a node g_i, in a graph network Υ, can be calculated as:

$$H_c(g_i) = \sum_{g_j \in \mathcal{G}} d(g_i, g_j) - \frac{\sum_{g_j \in \mathcal{G}} a_{g_i g_j} \times \sum_{k=1}^{|\mathcal{G}|} d(g_j, g_k)}{\sum_{g_j \in \mathcal{G}} a_{g_i g_j}}. \tag{14}$$

- **Degree** [4], which can serve as a good baseline for comparison.

To evaluate the lists of nodes ranked by all the centrality measures, we need to compare them with the ranked list that is obtained by the real spreading process of the nodes. Thus, we applied the following traditional schema [28]–[30] to validate the performance of a tested centrality measure:

1) Compute the centrality measure for all nodes, and then record the node ranking list
2) Use SIR model [28] to simulate the spreading ability of the nodes. In the SIR model, every node belongs to

one of the susceptible states: the infected state or the recovered state. At each step, we set only one node to be infected, the other nodes are susceptible nodes, and then investigate the information spreads in the network. Every infected node can infect its susceptible neighbours with spreading (also called infection) probability. Note that instead of considering the recovered state of each node, we focus on the influence within a time $t = 10$ since the spreading in early stage is found to be more important in practice. At the end of the SIR simulation process, we calculate the spreading efficiency for every node, and then record the node influence ranked list

3) Based on the centrality-based ranking list and the one generated by the SIR model, we record the joint score list $\mathcal{L} = \{(x_i, y_i)\}_{i=1}^{n}$, where x_i and y_i are the centrality-based and SIR-based measures of a node $g_i \in \mathcal{G}$, respectively. For any two randomly selected pairs $(x_i, y_i), (x_j, y_j) \in \mathcal{L}$, if both $(x_i < x_j)$ and $(y_i < y_j)$ or if both $(x_i > x_j)$ and $(y_i > y_j)$, they are said to be *concordant*. If both $(x_i < x_j)$ and $(y_i > y_j)$ or if both $(x_i > x_j)$ and $(y_i < y_j)$, they are said to be *discordant*. If $(x_i = x_j)$ and $(y_i = y_j)$, then the pair is neither concordant nor discordant.

Consequently, we considered the following two metrics to assess the accuracy and the scalability of the results:

1) The Kendall's tau rank correlation coefficient τ:

$$\tau = \frac{2(n_c - n_d)}{n(n-1)}, \tag{15}$$

where n_c and n_d are the number of concordant and discordant pairs in $\mathcal{L}$, respectively. A high τ value indicates that the centrality measure could produce an accurate ranked list. The ideal case is when $\tau = 1$ where the ranked list generated by the centrality measure is symmetrical to the ranked list generated by the real spreading process.

2) The average elapsed time ξ:

$$\xi = \frac{\sum_{g_i \in \mathcal{G}} t_i}{n}, \tag{16}$$

where t_i is the elapsed times for computing the underlying centrality measure of a node $g_i \in \mathcal{G}$.

All the experiments were run on an Intel(R) Core-i7 CPU @ 2.6GHz computer with 16 GB of memory under macOS Mojave. We implemented all considered indices as an extension to NetworkX Python package. For extracting the formal concepts from the lattice, we make use of the *Concepts 0.7.11* Python package, which is implemented by Sebastian Bank.[5]

C. RESULTS

We conducted our experimental evaluations through three experiments.

[5]Publicly available: https://pypi.python.org/pypi/concepts

Experiment I: The first experiment was dedicated to answering Q1. In the SIR model simulation, each infected node can infect its susceptible neighbours with a spreading probability β. Thus, in line with the scheme explained above, we repeatedly computed the joint list $\mathcal{L}$ of each centrality measure and the real spreading of the nodes while increasing the spreading probability β in the range $(0, 0.1]$ with increments of 0.01. On that basis, at each increment step, we calculated the corresponding evaluation metric τ in Eq. (15).

Figure 3 displays The Kendall's tau correlation coefficient τ between the the seven tested centrality measures and the ranking list generated by the SIR model, with a spreading probability $\beta \in (0, 0.1]$ and at a given time $t = 10$. Overall Cross-face outperforms all the centrality measures compared, achieving the most accurate Kendall coefficient τ on Netscience, USAir97 and Jazz musicians networks. For the Email network, Cross-face has the highest τ value when the spreading probability $\beta \leq 0.03$ or $\beta \geq 0.08$, otherwise Eigenvector slightly competes with Cross-face. The Eigenvector comes close behind Cross-face on USAir97, but considerably further behind on both Netscience and Jazz musicians networks. Heatmap is clearly more accurate than closeness on all the tested networks. K-shell is more accurate than degree on Netscience, USAir97 and Email networks, but it is outperformed by degree on Jazz musicians network. Closeness outperforms betweenness on both USAir97 and Jazz musicians networks, and on the contrary, betweenness is better than it on Email and Netscience networks when $\beta \geq 0.07$ or $\beta \leq 0.03$, respectively. Eigenvector is more accurate than both closeness and betweenness on Email, USAir97 and Jazz musicians networks when $\beta \leq 0.03$. On both Netscience and USAir97 networks, the Heatmap outperforms K-shell, but K-shell is marginally more accurate than it on both Email and Jazz musicians networks when $\beta < 0.03$. Remarkably, all the centrality measures are positively correlated at the whole range of spreading probability.

Experiment II: This experiment was performed to answer Q2. We are interested here in assessing the performance of the centrality measures. That is, we reran Experiment I while reporting their computational time as in Eq. 16. Figure 4 shows the average elapsed time ξ of the seven centrality measures on the four underlying networks. Overall, the cross-face dominates all other centrality measures on all networks tested. It finishes eight times faster than the betweenness, five times faster than both closeness and heatmap, and at least two times quicker than k-shell on Email, USAir97 and Jazz musicians networks. It also clearly prevailed over the eigenvector by a significant margin on all four networks. Degree is very competitive with cross-face on Email, Netscience and Jazz musicians networks, but cross-face is 1.25 times faster than it on USAir97 network. Apart from cross-face, and on all networks, the eigenvector is marginally faster than the closeness by at least a factor of 1.75. In addition, the closeness is considerably faster than betweenness, and competes with heatmap on Emails and Netscience networks. Eigenvector

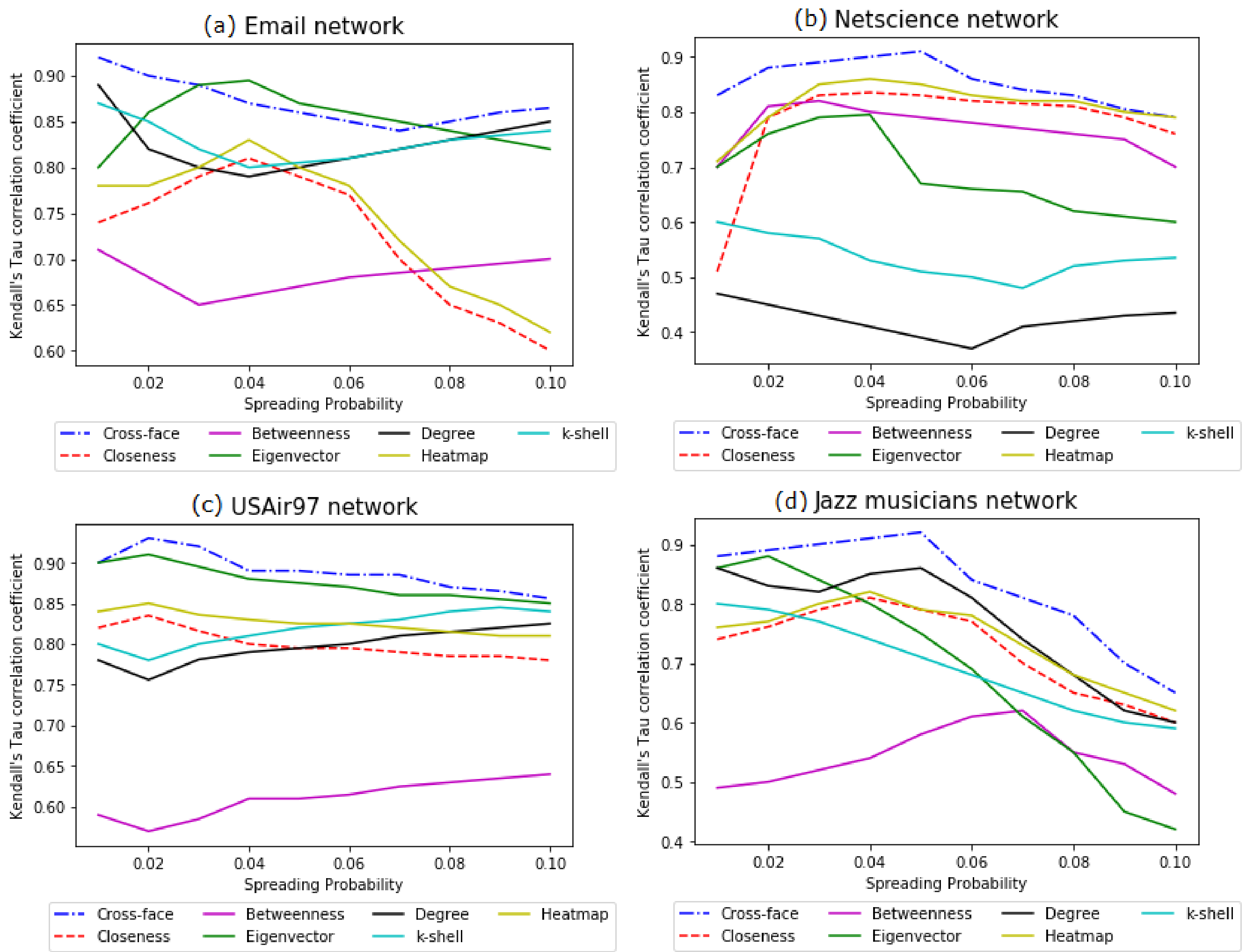

FIGURE 3. The Kendall's tau coefficient τ between the tested centrality measures and the ranking list generated by the SIR model, with $\beta \in (0, 0.1]$, at $t = 10$ on the four underlying datasets: (a) Email, (b) Netscience, (c) USAir97 and (d) Jazz musicians.

is significantly more faster than k-shell on Emails and Jazz musician networks, but on the contrary, k-shell is quicker than it on Netscience network.

Experiment III: Here we concentrate on Q3. The goal was to compare the most key nodes obtained using different centrality measures. Table 3 records the top-5 nodes of each measure in four networks. We use a symbol "$g_i \sqsupset g_j$" to denote that "node g_i is more important than node g_j". Succinctly, Cross-face identifies more accurate key nodes than all other tested centrality measures. On Email network, the cross-face indicated that 105 $\sqsupset$ 333 $\sqsupset$ 23 $\sqsupset$ 16 $\sqsupset$ 76. Degree coincides with the cross-face that the two nodes 105 $\sqsupset$ 333 are the two most key ones, whether Eigenvector matches it on that 105 is the top-1 node. Closeness and betweenness coincide with cross-face for only 333 $\sqsupset$ 23. Both of K-shell and heatmap behave differently from cross-face. As for Netscience network, the cross-face articulates that 78 $\sqsupset$ 281 $\sqsupset$ 150, which is consistent with closeness and heatmap, and partially matches betweenness on that 78 $\sqsupset$ 150, but contrary to eigenvector and degree. In USAir97 network, the cross-face elucidates that 118 $\sqsupset$ 261 $\sqsupset$ 255, which is identical to eigenvector and degree, and partially matches with betweenness, k-shell and heatmap. For Jazz musicians network, the top-1 node 60 is identified by the cross-face, eigenvector and k-shell, while the other four centrality detect 136 as the top-1 node. We can also see that the cross-face and eigenvector identify the nodes 136 and 132 as next less influential nodes. This partially coincides with degree and k-shell on the node 132, and is contrary to betweenness, closeness and heatmap.

D. DISCUSSION

In terms of accurate node centrality, the results of Experiment I, in Subsection IV-C, suggest that Cross-face outperforms traditional centrality measures such as degree, closeness, betweenness, k-shell and eigenvector. It improves the identification of highly central or topologically important nodes. This is attributable to the virtue of concurrently

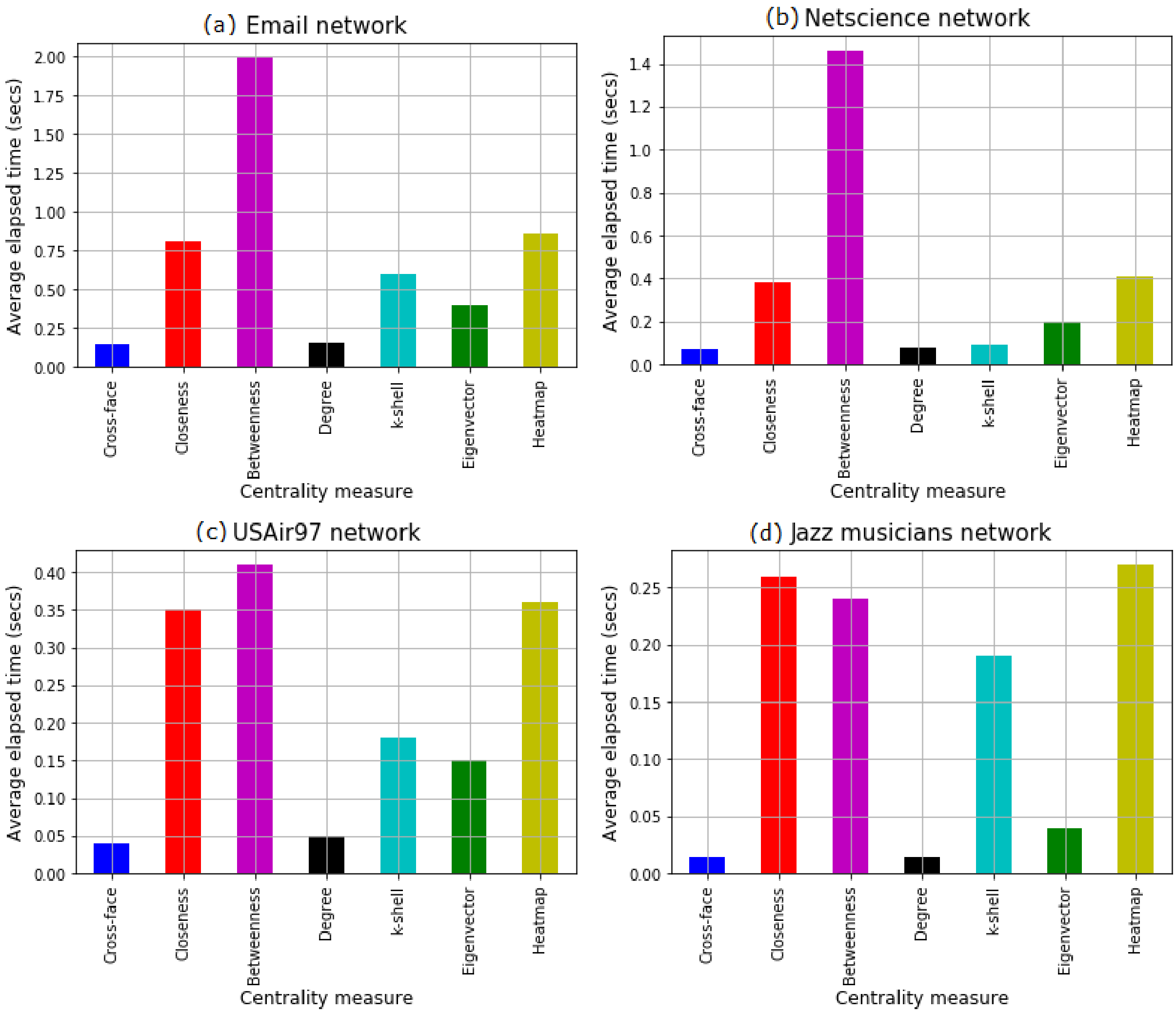

FIGURE 4. **Average elapsed time (in secs) ξ of the seven tested centrality measures: cross-face, closeness, betweenness, degree, eigenvector, k-shell and heatmap on (a) Email, (b) Netscience, (c) USAir97 and (d) Jazz musicians datasets.**

considering local and global aspects of network topology using its cross-clique and face-bridge terms, respectively. The cross-clique quantifies the structural embeddedness of dense regions in a network that involve the node. From a conceptual viewpoint, it captures the local information on how the node influences its immediate important nodal neighbours through the lens of adjacent cross-cliques. The face bridge term quantifies the global role of the node based on the routing of information along key bridges (i.e., important geodesics).

From the performance perspective, the results of Experiment II from the previous section, indicate that the cross-face is relatively faster than all other tested centrality measures. In practice, this is due to the fact that cross-face mainly computes the centrality of all nodes based on the set of symmetrical concepts $\tilde{\mathcal{C}}$, which is frequently quite small compared to the sets of nodes and edges that are used by all other tested centrality, i.e., $n_{\tilde{c}} \ll n$ and $n_{\tilde{c}} \ll m$.

Taking the correlation between centrality measures into consideration, the results of Experiment III, in Subsection IV-C, elucidate that cross-face centrality provides unique or correlated node identification based on the topology of the network. When the network contains a large number of dense regions (e.g., having a high average clustering coefficient) with many nodes having high degrees (e.g., having a high average degree) and there is a small number of hole structures, the role of the cross-clique term dominates the face-bridge one, and here the cross-centrality could be partially correlated with both degree and eigenvector centrality measures. This is because the network tends to decompose into several cohesive clusters (or communities), on which the nodes that have high degree are potentially the central ones. On the flip side of the coin, when the network contains a small number of (semi-symmetrical) dense regions, the role of the face-bridge term dominates the cross-clique one, even in the existence of a small number of structural

TABLE 3. The top-5 nodes ranking using Cross-face centrality (CF_c), Degree centrality (D_c), Betweenness centrality (B_c), Closeness centrality (C_c), Eigenvector centrality (EV_c), k-shell centrality (Ks_c) and Heatmap centrality (H_c) on the tested networks.

Network	Rank	CF_c	B_c	C_c	EV_c	D_c	Ks_c	H_c
Email	1	105	333	333	105	105	299	512
	2	333	105	23	16	333	389	23
	3	23	23	105	196	16	434	100
	4	16	578	42	204	23	552	219
	5	76	76	41	42	42	571	41
Netscience	1	78	78	78	1429	33	78	78
	2	281	150	281	1430	34	54	281
	3	150	516	150	1431	78	33	150
	4	516	281	756	645	54	150	301
	5	301	216	301	1432	294	756	756
USAir97	1	118	118	118	118	118	118	118
	2	261	8	261	261	261	261	261
	3	255	261	67	255	255	67	100
	4	201	201	255	182	152	201	255
	5	152	47	201	152	182	230	201
Jazz	1	60	136	136	60	136	60	136
	2	136	153	60	132	60	132	60
	3	132	60	168	136	132	168	149
	4	168	149	70	168	168	33	70
	5	108	168	83	108	70	100	83

holes. This is because the effect of bridges comes to play for determining the central nodes, and the cross-centrality could be slightly correlated with betweenness and closeness. For more clarification, let us consider the toy graph in Figure 5 as an illustrative example.

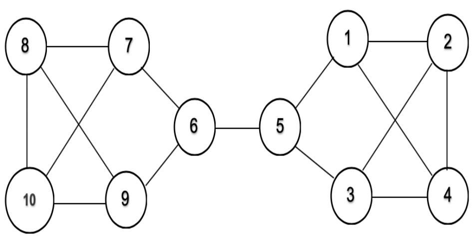

FIGURE 5. A toy graph network.

As shown in Table 4, both of degree and eigenvector can not identify which nodes are influential. The closeness and betweenness have the nodes {5, 6} as the most important, the nodes {1, 3, 7, 9} as the next most important, and {2, 4, 8, 10} as the least important. Similarly, the Cross-face centrality identifies the nodes {5, 6} as the most influential ones, but on the contrary to betweenness and closeness, it ranks the nodes {2, 4, 8, 10} as the next most important, and {1, 3, 7, 9} as the least influential ones.

TABLE 4. Cross-face centrality (CF_c), Degree centrality (D_c), Betweenness centrality(B_c), Closeness centrality (C_c), Eigenvector centrality(EV_c) centrality measures of the toy graph in Figure 5.

Centrality		D_c	C_c	B_c	EV_c	CF_c
Node	**1**	0.33	42.9	0.167	0.32	0.225
	2	0.33	34.6	0.009	0.32	*0.25*
	3	0.33	42.9	0.167	0.32	0.225
	4	0.33	34.6	0.009	0.32	*0.25*
	5	0.33	**52.9**	**0.565**	0.32	**0.30**
	6	0.33	**52.9**	**0.565**	0.32	**0.30**
	7	0.33	42.9	0.167	0.32	0.225
	8	0.33	34.6	0.009	0.32	*0.25*
	9	0.33	42.9	0.167	0.32	0.225
	10	0.33	34.6	0.009	0.32	*0.25*

In other topological scenarios of the network such as a mixture of dense regions and structural holes, both of the cross-clique and face-bridge terms could approximately have the same merits, and here the cross-face centrality is anticipated to produce unique results. Apart from Cross-face centrality, it is generally observed that the conceptual distinctions between other centrality measures do not frequently appear as empirical differences in networks. This means that two centrality indices could have distinct theoretical foundations while showing practically redundant results since they behave similarly on a given network.

V. CONCLUSION AND FUTURE WORK

The identification of key nodes in complex networks is a key step in the development of scientific data mining systems. We believe that there is a clear gap in the existing complex network analysis literature on how to efficiently exploit FCA formulations to encapsulate both the cohesion and the shortest path based centrality concepts in one measure. On that basis, we proposed *Cross-face*, a new FCA-based centrality measure to quantify the importance of a given node within the network formulation based on its presence in actionable cross cliques and bridges.

The novelty of the cross-face centrality framework lies in the following: first, we provide a concept lattice formulation of the network that facilitates the efficient extraction of its cliques and bridges through the so-called symmetrical concepts; second, we leverage the faces of those symmetrical concepts to identify key bridges; third, FCA centrality measurement is defined to quantify the importance of nodes within the network based on how it influences other nodes by hybridizing two different but complementary aspects: cross-cliques and bridges. The thorough empirical study on several real-life networks (see Section IV) shows that the cross-face centrality can quantify the importance of nodes in a more accurate and efficient manner than other state-of-the-art centrality indices like degree, betweenness, closeness, k-shell and eigenvector.

There are still a number of points for future work. We plan to generalize the Cross-face centrality to detect key nodes in *two-mode* data and *multi-layer* networks and *multi-dimensional* ones. We also intend to propose an online variant of Cross-face algorithm applicable to dynamic networks. Finally, since the computation of the cross-clique and face-bridge terms are independent, we will design a high-speed cross-centrality by parallelizing the calculations of Eq. (13), and then intensively investigate its efficiency on more big, dense and complex data networks.

REFERENCES

[1] M. O. Jackson, *Social and Economic Networks*. Princeton, NJ, USA: Princeton Univ. Press, 2010.

[2] 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.

[3] S. Oldham, B. Fulcher, L. Parkes, A. Arnatkevic̆iūtė, 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.

[4] S. Tsugawa and H. Ohsaki, "Analysis of the robustness of degree centrality against random errors in graphs," in *Complex Networks VI*. New York, NY, USA: Springer, 2015, pp. 25–36.

[5] 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.

[6] M. R. Faghani and U. T. Nguyen, "A study of XSS worm propagation and detection mechanisms in online social networks," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 11, pp. 1815–1826, Nov. 2013.

[7] M. G. Everett and S. P. Borgatti, "Analyzing clique overlap," *Connections*, vol. 21, no. 1, pp. 49–61, 1998.

[8] N. Chiba and T. Nishizeki, "Arboricity and subgraph listing algorithms," *SIAM J. Comput.*, vol. 14, no. 1, pp. 210–223, Feb. 1985.

[9] U. Brandes, "A faster algorithm for betweenness centrality," *J. Math. Sociol.*, vol. 25, no. 2, pp. 163–177, 2001.

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

[11] M. E. Newman, "The mathematics of networks," in *The New Palgrave Encyclopedia of Economics*, vol. 2. London, U.K.: Palgrave Macmillan, 2008, pp. 1–12.

[12] T. Martin, X. Zhang, and M. E. J. Newman, "Localization and centrality in networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 90, no. 5, Nov. 2014, Art. no. 052808.

[13] D. Easley and J. Kleinberg, *Networks, Crowds, and Markets*, vol. 8. Cambridge, U.K.: Cambridge Univ. Press, 2010.

[14] M. Granovetter, "The strength of weak ties: A network theory revisited," *Sociol. Theory*, vol. 1, no. 1, pp. 201–233, 1983.

[15] B. Ganter and R. Wille, *Formal Concept Analysis: Mathematical Foundations*. New York, NY, USA: Springer-Verlag, 1999.

[16] P. Valtchev, R. Missaoui, and P. Lebrun, "A partition-based approach towards constructing Galois (concept) lattices," *Discrete Math.*, vol. 256, no. 3, pp. 801–829, Oct. 2002.

[17] V. Choi, "Faster algorithms for constructing a concept (Galois) lattice," in *Clustering Challenges in Biological Networks*. Singapore: World Scientific, 2009, pp. 169–186.

[18] J. L. Pfaltz and C. M. Taylor, "Scientific knowledge discovery through iterative transformation of concept lattices," in *Proc. SIAM Workshop Discrete Math. Data Mining*, Arlington, VA, USA, 2002, pp. 1–10.

[19] K. Okamoto, W. Chen, and X.-Y. Li, "Ranking of closeness centrality for large-scale social networks," in *Proc. Int. Workshop Frontiers Algorithmics*. Changsha, China: Springer, 2008, pp. 186–195.

[20] M.-H. Ibrahim, R. Missaoui, and A. Messaoudi, "Detecting communities in social networks using concept interestingness," in *Proc. 28th Annu. Int. Conf. Comput. Sci. Softw. Eng.*, 2018, pp. 81–90.

[21] R. Guimera, L. Danon, A. Díaz-Guilera, F. Giralt, and A. Arenas, "Self-similar community structure in a network of human interactions," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 68, no. 6, Dec. 2003, Art. no. 065103.

[22] M. E. J. Newman, "Finding community structure in networks using the eigenvectors of matrices," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 74, no. 3, Sep. 2006, Art. no. 036104.

[23] R. A. Rossi and N. K. Ahmed, "The network data repository with interactive graph analytics and visualization," in *Proc. 29th AAAI Conf. Artif. Intell.*, 2015, pp. 1–2. [Online]. Available: http://networkrepository.com

[24] V. Krebs. (2008). *A Network of Books About Recent us Politics Sold by the Online Bookseller Amazon.Com*. [Online]. Available: http://www.orgnet.com

[25] Q. Hu, Y. Gao, P. Ma, Y. Yin, Y. Zhang, and C. Xing, "A new approach to identify influential spreaders in complex networks," in *Proc. Int. Conf. Web-Age Inf. Manage.* Beidaihe, China: Springer, 2013, pp. 99–104.

[26] J. Xu, "KSC centralized index model in complex network," *J. Netw.*, vol. 9, no. 5, p. 1245, May 2014.

[27] C. Durón, "Heatmap centrality: A new measure to identify super-spreader nodes in scale-free networks," *PLoS ONE*, vol. 15, no. 7, Jul. 2020, Art. no. e0235690.

[28] D. Chen, L. Lü, M.-S. Shang, Y.-C. Zhang, and T. Zhou, "Identifying influential nodes in complex networks," *Phys. A, Statist. Mech. Appl.*, vol. 391, pp. 1777–1787, Feb. 2012.

[29] N. Zhao, S. Fang, N. Chen, and C. Pei, "Modeling and analyzing the influence of multi-information coexistence on attention," *IEEE Access*, vol. 7, pp. 117152–117164, 2019.

[30] N. Zhao, J. Bao, and N. Chen, "Ranking influential nodes in complex networks with information entropy method," *Complexity*, vol. 2020, pp. 1–15, Jun. 2020.

MOHAMED HAMZA IBRAHIM received the Ph.D. degree in computer engineering from the University of Montreal, Canada, in 2015. He is currently a Postdoctoral Researcher with the University of Quebec in Outaouais, Canada. He is also a Lecturer of artificial intelligence with the Department of Mathematics, Faculty of Science, Zagazig University, Egypt. Previously, he worked as a Postdoctoral Researcher with the Polytechnique Montreal, University of Montreal, Canada. He also worked as a Data Scientist with the Royal's Institute of Mental Health Research, Canada. His research interests include deep/machine learning, big data mining, formal concept analysis, social network analysis, and constraint programming.

ROKIA MISSAOUI received the Ph.D. degree in computer science from the University of Montreal, in 1988. She has been a University Professor in Canada for more than 33 years. She is currently a Full Professor with the Department of Computer Science and Engineering, University of Quebec in Outaouais (UQO). Before joining UQO in 2002, she was a Professor with the University of Quebec in Montreal (UQAM) for 15 years. She leads the LARIM Laboratory, UQO. She is also a member of the LATECE Laboratory, UQAM. Since the beginning of her career, she has been involved in several research projects funded by granting agencies and industrial partners. Her current research interests include advanced databases, data mining and warehousing, machine learning, social network analysis, and big data analytics.

JEAN VAILLANCOURT received the Ph.D. degree in mathematics from Carleton University, Ottawa, ON, Canada, in 1987. He is currently a Professor with HEC Montreal, after two full mandates as the President of the Université du Québec en Outaouais. He has served on numerous committees, notably as a member of the boards of directors of the Statistical Society of Canada from 1997 to 2001 and the Institute of Materials and Advanced Mechanics in Le Mans, France, from 2005 to 2015, and the Chair of the Mathematics and Statistics Group, Natural Sciences and Engineering Research Council of Canada, from 2003 to 2006. His current research interests include theoretical studies of random media as well as knowledge discovery in large databases.