Received November 5, 2021, accepted November 22, 2021, date of publication November 30, 2021,
date of current version December 8, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3131987

Identifying Influential Nodes in Two-Mode Data Networks Using Formal Concept Analysis

MOHAMED HAMZA IBRAHIM[1,2], ROKIA MISSAOUI[1], AND JEAN VAILLANCOURT[3]
[1]Department of CS and Engineering, University of Quebec in Outaouais, Gatineau, QC J8Y 3G5, Canada
[2]Department of Mathematics, Faculty of Science, Zagazig University, Zagazig 44519, Egypt
[3]Department of Decision Sciences, HEC Montréal, Montréal, QC H3T 2A7, Canada

Corresponding author: Mohamed Hamza Ibrahim (mohamed.ibrahim@polymtl.ca)

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

ABSTRACT Identifying important actors (or nodes) in a two-mode network is a crucial challenge in mining, analyzing, and interpreting real-world networks. While traditional bipartite centrality indices are often used to recognize key nodes that influence the network information flow, inaccurate results are frequently obtained in intricate situations such as massive networks with complex local structures or a lack of complete knowledge about the network topology and certain properties. In this paper, we introduce *Bi-face* (BF), a new bipartite centrality measurement for identifying important nodes in two-mode networks. Using the powerful mathematical formalism of Formal Concept Analysis, the BF measure exploits the faces of concept intents to detect nodes that have influential bicliques connectivity and are not located in irrelevant bridges. Unlike off-the shelf centrality indices, it quantifies how a node has a cohesive substructure influence on its neighbour nodes via bicliques while not being in network core-peripheral ones through its absence from non-influential bridges. In terms of identifying accurate node centrality, our experiments on a variety of real-world and synthetic networks show that BF outperforms several state-of-the art bipartite centrality measures, producing the most accurate Kendall coefficient. It provides unique node identification based on network topology. The findings also demonstrate that the presence of terminal nodes, influential bridges, and overlapping key bicliques impacts both the performance and behaviour of BF as well as its relationship with other traditional centrality measures. On the datasets tested, the computation of BF is at least twenty-three times faster than betweenness, eleven times faster than percolation, nine times faster than eigenvector, and ten times faster than closeness.

INDEX TERMS Formal concept analysis, two-mode networks, influential node, cross-clique connectivity.

I. INTRODUCTION

In today's world, complex real-life systems are ubiquitous. For example, mobile phone as well as Facebook and Twitter networks facilitate the way we interact with others. The energy and electric power networks play a significant role in supplying our domestic and industrial lives. Most of these systems frequently feature two types of data with complex substructures and can thus be represented as two-mode networks (also known as bipartite graphs or affiliation networks). Due to the complex structure of such networks, the spread of information across the network makes some nodes more important than others in certain contexts. Our main motivation here is to solve the problem of identifying key actors in such networks. This is the case for a wide range of real-world applications, such as the identification of the most active agents who serve customers in call centres at various locations, the most popular products in various hot-spot regions, the most cited authors collaborating on scientific papers published in top-tier conferences, the actors who have a major influence on audiences in different movies, or proteins that have a large impact on various metabolic processes. As such, the interesting question of how to measure the relative importance of nodes in a two-mode network is often increasingly challenging in the field of complex network analysis (CNA). As it is frequently used to understand the role of nodes within a network, node centrality analysis can provide efficient answers to this question.

The centrality measure ranks nodes based on how they influence or are effected by other nodes via their

The associate editor coordinating the review of this manuscript and approving it for publication was Jolanta Mizera-Pietraszko.

connection topology. Since no consensus holds on a unique definition of centrality for two-mode networks in the CNA literature, while opening the door for the invention of new ones, different centrality scores have been presented (cf. [1], [2] for a detailed survey), each of which takes into account a distinct aspect of a central node. In the mainstream CNA research area, the bipartite centrality is frequently classified as local or global. The local centrality metrics focus on the relative importance of a node in its neighbourhood within local cohesive communities. For example, the degree centrality [3] is a basic local metric that counts the number of links that each node has. However, *it frequently captures irrelevant local information about a node in practice*. Intuitively, it is assumed that only the node with the highest degree should be in the centre (because it is the most densely linked node i.e., a hub), but it does not account for the cascade effects of its neighbour nodes. Hence, it is sometimes necessary to remove nodes with high degree values because they provide no information. For example, Angelina Jolie has a high degree centrality in Facebook's network because so many people follow her. However, if you explore your friends' Facebook pages to find out what they are interested in or who among them enjoy soccer the most, Angelina Jolie becomes completely irrelevant in that network.

The k-shell centrality [4] is a community-based local centrality that enhances the degree of a node in terms of its neighbourhood connections using the k-core.[1] Thus, the higher the portions of k-cores contain a node, the more likely it is a hub in the cores of a network, and thus the more important it is in a network. However, *k-shell frequently produces inaccurate results when the network structure has a small number of k-cores, which is prevalent in two-mode networks*. This is due to the fact that in this case, many nodes are assigned an approximately equal number of k-cores. From the perspective of the topological graph of a two-mode network, k-bicliques may be more accurate graphical components than k-cores. That is, the number of k-bicliques among a node neighbours is counted in order to estimate its importance using the Cross k-bicliques connectivity measure, which quantifies how the node affects information propagation through the network. However, in general, *its calculation requires an exponential time and space complexity and is often sensitive to the k parameter*. To compute Cross k-bicliques connectivity for a given node, we must first extract all k-bicliques from the network containing this node, which is an NP-hard problem. Furthermore, the determination of the optimal value of k may be problematic in many applications. Strictly speaking, picking a large k value may result in the overstepping of all k-bicliques with k less than the chosen one, leading to an underestimation of the influence of other nodes in local cohesive communities within the network. A small k value may stimulate overestimation of the importance of other neighbour nodes, generating a behaviour similar to degree centrality.

Bipartite closeness [3], [5] is a common type of local centrality that is based on the geodesics. It computes the reciprocal of the sum of the distances between the node and all of the other nodes in the network. Its basic form intuitively assumes that information can efficiently flow from one node to every other node via the shortest distances. The important node is therefore the independent one that is close to other nodes in the network in terms of shortest paths. Thus, at a high level, it can address the degree centrality limitation in a few cases. However, *on non-spatial networks, bipartite closeness frequently produces inaccurate results [6], and its values on spatial networks tend to span a rather small dynamic range from the smallest to the largest*. This is because most complex real-world networks may have a high average length of the shortest path as their largest distance increases exponentially in terms of the number of nodes. That is, assuming that the minimum distance is equal to one, the asymptotic ratio between the minimum and the largest distances is $O(\frac{1}{\log n})$. This frequently implies that numerous nodes, with diverse roles in the network information flow, may have comparable closeness scores. On the contrary, most non-spatial networks feature low geodesic distances among nodes given that high geodesic distances increase logarithmically with their network size. As a result, the dynamic range of variations, as well as the network diameter, will be too small, and even slight changes in the network structure can have a significant impact on nodal closeness values.

The bipartite betweenness [3], [5], [7] is another common geodesic-based measure. To evaluate the importance of a node, it computes the number of times it exists in the bridge along the geodesic paths among the other nodes in the network. Thus, it considers other nodes' dependence on a given node, and measures its optimal flow control on information passing among nodes, whether closeness perceives the connection efficiency or independence from potential flow control through the use of intermediary nodes (cf. [8], a detailed study differentiating between closeness and betweenness). In general, *bipartite betweenness does not consider node connectivity and its calculation is frequently time-consuming*. The fundamental assumption of betweenness is that every pair of nodes exchanges information through shortest-paths with equal probability. However, this is, in many situations, not a realistic assumption since information does not necessarily take the shortest path [9] (e.g., news related to a friend might not be directly known from another close friend but from other mutual friends). As a result, it does not provide a precise representation of the most influential nodes within these groups, but rather a fair approximation (see [10] for a more detailed explanation). Furthermore, its exact centrality computation on large or dense two-mode networks requires a time complexity of $O(n_1^3 + n_2^3)$, where n_1 and n_2 are the number of the two types of nodes, respectively.

Looking at local centrality from a different angle, bipartite percolation centrality [11] estimates a node's relative

[1] A k-core of a graph G is a maximal connected sub-graph of G such that all nodes have at least k neighbours.

importance by counting the number of percolated paths that pass through it. The percolated path is the shortest path between two nodes in which the source node is percolated (e.g., infected) but the target node may not be. The percolation centrality fully captures the essential mechanics of contagion-mediated network spreading by associating percolation paths with weight terms that determines how much importance is given to potential percolation paths originating from given nodes. This is indeed helps percolation centrality to avoid the limitation of both betweenness and closeness, which rely solely on topological and random diffusion processes via random shortest-paths. *It may, however, produce inaccurate results when the spread of contagion has no effect on changing the node state, and it is frequently computationally expensive to calculate*. Because the percolation through a network is affected by both the level of contagion and the network structure [12], the spread of contagion in a complex network (CN) may not change node states in a few scenarios. From a theoretical standpoint, it is possible that there is no transmissibility, and in this case, the percolated contagion spreads over the edges of a complex network without changing the state of a node to either recoverable or infected, leaving it in the default state. Moreover, computing the percolation centrality in worst-case scenarios with large bipartite networks having complex local structures requires a cubic time complexity in the two types node numbers.

Global measures, on the other hand, consider a node prominence in the context of the entire network. Its principle emphasizes the hypothesis that a few important neighbours can weight more than a large number of unimportant ones. That is, a node is important if it is connected to other important nodes. For example, Bipartite Eigenvector centrality [3], [5] quantifies whether a node is central based on its connections to other high-score nodes. It utilizes indefinite-length random walks to estimate the number of node traversals. From a conceptual standpoint, a node's eigenvector can be thought of as the global extension of its local degree centrality, in which both count walks that begin and terminate at that node. *Eigenvector may include a localization transition, which frequently results in inaccurate centrality scores*. As demonstrated in [13], eigenvector centrality has a localization transition under the common conditions of a network regime, causing the majority of the weight of the centrality to concentrate on a small number of nodes in the network. This implies that when a network structure contains many hubs, the eigenvector weights are skewed toward some few nodes: The eigenvector values of the hub node and its neighbours are the highest, while the other nodes have identical centrality values (likely close to zero).

In this paper, we present **Bi-face** (BF), a new bipartite centrality that can be used to identify key nodes in two-mode networks. From the standpoint of graph modelling, the superiority of the bipartite graph formulation of two-mode networks over other graphical models such as concept lattices, factor graphs, or graph neural networks is still uncertain. That is, it is dependent on the task we aim to apply to the problem. One of our goals here is to show that the concept lattice model of such problems is more advantageous than bipartite graph formulation for efficiently computing two-mode node centrality. Thus, the guiding idea of BF is to use a formal concept analysis framework to bring together the centrality aspects of cohesiveness via bicliques, network flow via bridges, and influence of important neighbour nodes for the benefit of actionable node identification. Its conceptual hypothesis is based on the fact that important nodes should be found in influential bridges and overlapping bicliques with a large number of important nodes. That is, it quantifies how a node affects, and is affected by, its important neighbours via bicliques while also connecting the densely substructures of a network through its presence in influential bridges. Thus, it differs from betweenness in that it deems influential bridges rather than all bridges. Unlike closeness and eigenvector, it can efficiently deal with the diverse topological structures of a network, without potentially having localization transition, due to this hybridization of the influential bridges and overlapping bicliques aspects. Furthermore, it leverages the powerful mathematical formulation of Formal Concept Analysis (FCA) to overcome the limitation of Cross k-bicliques connectivity. That is to say, it utilizes the concept lattice related to the network to efficiently extract concepts that capture bridges and k-bicliques from the network while being insensitive to the k parameter. Technically, BF computation is based solely on the set of these extracted concepts, which is often quite small in comparison with polynomial functions in terms of nodes and edges. As a result, in contrast to percolation, it is relatively quick to compute in practice.

The paper is organized as follows. In Section II, we review some basic definitions and concepts including: FCA and traditional bipartite centrality measures in social networks. In Section III, we demonstrate our proposed Bi-face centrality for detecting influential nodes of two-mode networks in further more detail. In Section IV we conduct a thorough experimental study and a discussion. Finally, Section V presents our conclusions.

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 a two-mode network of airline companies and their flying destinations in the year 2000. As shown in Figure 1, the network is modelled as an undirected bipartite graph $\Upsilon = (\mathcal{G}, \mathcal{M}, \mathcal{I})$, where $\mathcal{G}$ is a set of 13 objects (also called type-I nodes) representing Star Alliance airline companies, $\mathcal{M}$ is a set of 9 attributes (type-II nodes) representing flying destinations, and $\mathcal{I}$ is a set of edges where an edge $(u_i, v_j) \in \mathcal{I}$ links two nodes $u_i \in \mathcal{G}$ and $v_j \in \mathcal{M}$, if a flight from airline company u_i lands at the destination v_j.

A. FORMAL CONCEPT ANALYSIS

In the following we recall notions of FCA [14] that will be used in this paper.

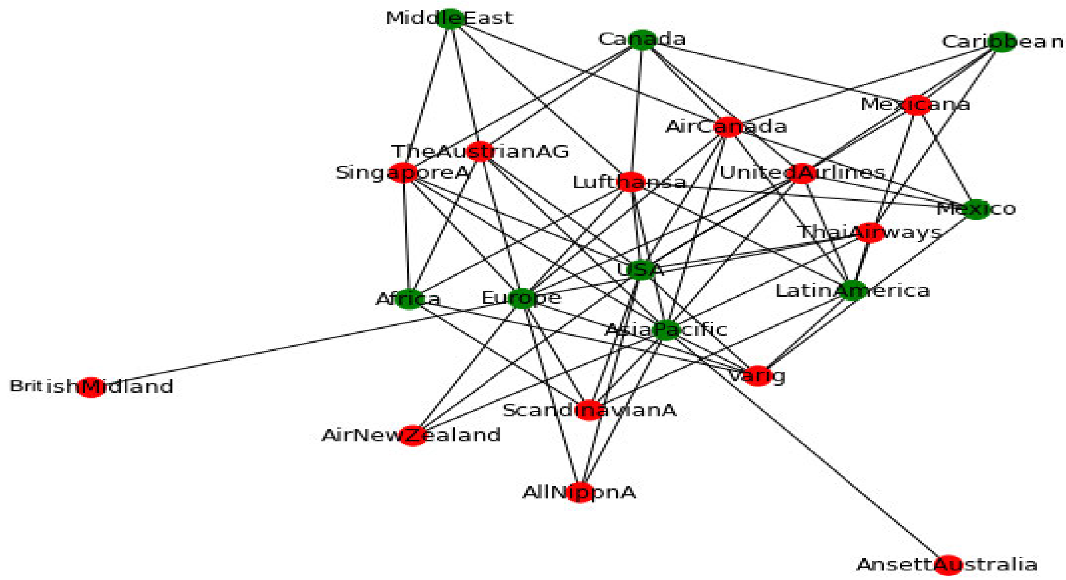


FIGURE 1. A two-mode graph network representing flights from 13 star alliance airline companies (in red) landing at 9 destinations (in green) in year 2000.

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

$\mathcal{G}$	LatinAmerica	Europe	Canada	AsiaPacific	MiddleEast	Africa	Mexico	Caribbean	USA
AirCanada	1	1	1	1	1	0	1	1	1
AirNewZealand	0	1	0	1	0	0	0	0	1
AllNippnA	0	1	0	1	0	0	0	0	1
AnsettAustralia	0	0	0	1	0	0	0	0	0
TheAustrianAG	0	1	1	1	1	1	0	0	1
BritishMidland	0	1	0	0	0	0	0	0	0
Lufthansa	1	1	1	1	1	1	1	0	1
Mexicana	1	0	1	0	0	0	1	1	1
ScandinavianA	1	1	0	1	0	1	0	0	1
SingaporeA	0	1	1	1	1	1	0	0	1
ThaiAirways	1	1	0	1	0	0	0	1	1
UnitedAirlines	1	1	1	1	0	0	1	1	1
Varig	1	1	0	1	0	1	1	0	1

Definition 1 (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., $(g, m) = 0$).

Table 1 is the formal context equivalent to the adjacency matrix that expresses the two-mode network exhibited in Figure 1.

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 2 (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 $B' = A$.

The object concept $g \in \mathcal{G}$ is expressed by $\gamma g := \left(g'', g'\right)$ and the attribute concept of $m \in \mathcal{M}$ is defined by $\mu m := \left(m', m''\right)$.

Definition 3 (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 4 (Concept Lattice): The concept lattice of a formal context $\mathbb{K}$, denoted by $\mathfrak{B}(\mathbb{K}) = (\mathcal{C}, \preceq)$, is a Hasse diagram that represents all formal concepts $\mathcal{C}$ together with the partial order that holds between them. In $\mathfrak{B}(\mathbb{K})$, each node represents a concept with its extent and intent while the edges represent the partial order between concepts.

Figure 3 is the Hasse diagram of the concept lattice that corresponds to the context of Table 1. More precisely, it is a diagram with reduced labeling. This means that the label g is written below γg and m above μm. The extent of a concept represented by a node a is given by all labels in $\mathcal{G}$ from the node a downwards, and the intent by all labels in $\mathcal{M}$ from a upwards. For example, the node indicated by the red arrow represents the formal concept whose extent contains Mexicana, ThaiAirways, UnitedAirlines, and AirCanada by collecting the object labels in white boxes from the current node downward to the lattice infimum, while its intent contains the attribute labels in grey boxes LatinAmerica, Caribbean, and USA collected from the current node upward to the lattice supremum. Here the current node whose label is Caribbean is one of the direct predecessor (lower cover) of the node named UnitedAirlines, and the direct successor (upper cover) of the nodes labeled by Mexicana and ThaiAirways.

Definition 5 (Formal Context): It is a formal context $\tilde{\mathbb{K}} = (\mathcal{G}, \mathcal{M}, \mathcal{I})$ in which $\mathcal{G}$ is set of objects and $\mathcal{G}$ is the set of attributes, and $\mathcal{I}$ is a set of relations defined on $\mathcal{G}$ and $\mathcal{M}$ with $\mathcal{I} \subseteq \mathcal{G} \times \mathcal{M}$. For $g_i \in \mathcal{G}$ and $g_j \in \mathcal{G}$, $(g_i, g_j) \in \mathcal{I}$ holds iff object g_i is linked to g_j.

There are several methods (cf. [14]–[16]) that build the lattice, i.e., compute all the concepts together with the partial order.

Definition 6 (Lower and Upper Covers): For any two formal concepts $c_1 = (A_1, B_1) \preceq c_2 = (A_2, B_2)$ if:

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

or

$$A_1 \subseteq A_3 \subseteq A_2 \iff B_1 \supseteq B_3 \supseteq B_2, \quad (3)$$

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 $\mathcal{U}(c)$ and $\mathcal{L}(c)$ to denote the sets of upper and lower covers of the formal concept c respectively.

Definition 7 (Concept Intentional Face [17]): The *intentional face* $f_{\text{in}}(c, c_d)$ of a concept $c = (A, B)$ w.r.t. its d-th upper cover concept, $c_d = (A_d, B_d) \in \mathcal{U}(c)$, is the difference between their intent sets as: $f_{\text{in}}(c, c_d) = B \setminus B_d$.

Definition 8 (Concept Extensional Face): The *extensional face* $f_{\text{ex}}(c, c_l)$ of a concept $c = (A, B)$ w.r.t. its l-th lower cover concept, $c_l = (A_l, B_l) \in \mathcal{L}(c)$, is the difference between their extent sets as: $f_{\text{ex}}(c, c_l) = A \setminus A_l$.

Definition 9 (Blocker [17]): Given the family of faces Λ_c, the set Z is said to be a *blocker* of Λ_c if $\forall f_i \in \Lambda_c,\ f_i \cap Z \neq \emptyset$, and the blocker Z is said to be *minimal* if $\nexists Z_j \subset Z,\ \forall f_i \in \Lambda_c,\ f_i \cap Z_j \neq \emptyset$.

Definition 10 (Generator [18]): Given a concept $c = (A, B)$ in a formal context $\mathbb{K} = (\mathcal{G}, \mathcal{M}, \mathcal{I})$, a subset $H \subseteq B$ is called a *generator* of c iff $H'' = B$, and it is a *minimal generator* when $\nexists H_1 \subseteq H$ such that $H_1'' = B$. We use $\mathcal{H}_c^{\text{ex}}$ and $\mathcal{H}_c^{\text{in}}$ to denote the sets of minimal generators of a concept c w.r.t. its extent and intent respectively.

For example, {*Canada*} is a generator associated with the intent {*Canada*, *USA*}, and allows us to infer that whenever an airline has a Canada destination, then it necessarily has an USA destination.

B. SOCIAL NETWORK ANALYSIS

Definition 11 (Biclique): Let $\Upsilon = (\mathcal{G}, \mathcal{M}, \mathcal{I})$ be an undirected bipartite graph defined over the objects $\mathcal{G}$ and attributes $\mathcal{M}$. A biclique $\tilde{Q} = (\tilde{\mathcal{G}}, \tilde{\mathcal{M}})$ is a complete subgraph of Υ induced by a pair of two disjoint subsets $\tilde{\mathcal{G}} \subseteq \mathcal{G}$, $\tilde{\mathcal{M}} \subseteq \mathcal{M}$, such that $\tilde{\mathcal{G}} \neq \emptyset$, $\tilde{\mathcal{M}} \neq \emptyset$, $\forall u \in \tilde{\mathcal{G}}$, $\forall v \in \tilde{\mathcal{M}}$, $(u, v) \in \mathcal{I}$. The disjoint subsets $\tilde{Q} =$ ({AirCanada, Mexicana, ThaiAirways,UnitedAirlines}, {LatinAmerica, Caribbean, USA}) is an example of a biclique. Henceforth, we use $\tilde{Q}$ as our illustrative biclique (see the lattice node indicated by a red arrow in Figure 3) to support the understanding of definitions and principles related to the Bi-face centrality.

Definition 12 (Bridge): An edge $(u, v) \in \mathcal{I}$ of a two-mode data network Υ 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 instance, the edge (AnsettAustralia, AsiaPacific) represents a bridge in Υ.

Definition 13 (Bipartite Centrality Measure): The *centrality measure* of a type-I node $u \in \mathcal{G}$ is a function that assigns a positive real number to u quantifying its centrality w.r.t. to all other type-II nodes $v \in \mathcal{M}$ in the network Υ (and vice versa).

In two-mode networks, bipartite (also known as two-mode) centrality measures are commonly utilized to detect important nodes. Although numerous centrality metrics have been proposed, the degree, closeness, betweenness, and eigenvector have been demonstrated to be the most outstanding in a variety of applications, and they are thus widely used.

Definition 14 (Degree Centrality D_c [3], [19]): The degree centrality of a node in a two-mode graph network Υ is defined as:

$$\mathrm{D_c}(u_i) = \sum_{v_j \in \mathcal{M}} I_{ij}, \quad \forall u_i \in \mathcal{G}, \quad (4)$$

$$\mathrm{D_c}(v_j) = \sum_{u_i \in \mathcal{G}} I_{ij}, \quad \forall v_j \in \mathcal{M} \quad (5)$$

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

Definition 15 (Closeness Centrality C_c [3], [5]): The normalized closeness centrality of a node g_i, in a two-mode graph network Υ, is defined as:

$$C_c(u_i) = \frac{|\mathcal{M}| + 2(|\mathcal{G}| - 1)}{\sum_{v_j \in \mathcal{M}} d(u_i, v_j)}, \quad \forall u_i \in \mathcal{G}, \tag{6}$$

$$C_c(v_j) = \frac{|\mathcal{G}| + 2(|\mathcal{M}| - 1)}{\sum_{u_i \in \mathcal{G}} d(u_i, v_j)}, \quad \forall v_j \in \mathcal{M} \tag{7}$$

where $d(u_i, v_j)$ is the geodesic distance (shortest path) between the nodes u_i and v_j.

Definition 16 (Betweenness Centrality B_c [7]): In bipartite networks Υ, the normalized betweenness centrality of a node is defined as in [5]:

$$B_c(u_i) = \sum_{u_j \neq u_k \neq u_i,\ u_j, u_k, u_i \in \mathcal{G}} \frac{\sigma_{u_j u_k}(u_i)}{\sigma_{u_j u_k}}, \quad \forall u_i \in \mathcal{G}, \tag{8}$$

$$B_c(v_j) = \sum_{v_j \neq v_k \neq v_i,\ v_j, v_k, v_i \in \mathcal{M}} \frac{\sigma_{v_i g_k}(v_j)}{\sigma_{v_i v_k}}, \quad \forall v_j \in \mathcal{M}, \tag{9}$$

where $\sigma_{x_j x_k}$ denotes the total number of shortest paths between nodes x_j and x_k, and $\sigma_{x_j x_k}(x_i)$ is the number of those paths that traverse g_i. To normalize the betweenness, we simply divide $B_c(u_i)$ and $B_c(v_j)$ by the corresponding term to its node set [5]:

$$B_c(\mathcal{G}) = \frac{1}{2}\big[|\mathcal{M}|^2(s+1)^2 + |\mathcal{M}|(s+1)(2t-s-1) - t(2s-t+3)\big], \quad \forall u_i \in \mathcal{G}, \tag{10}$$

where $s = (|\mathcal{G} - 1| \operatorname{div} |\mathcal{M}|)$ and $t = (|\mathcal{G} - 1| \mod |\mathcal{M}|)$,

$$B_c(\mathcal{M}) = \frac{1}{2}\big[|\mathcal{G}|^2(p+1)^2 + |\mathcal{G}|(p+1)(2r-p-1) - r(2p-p+3)\big], \quad \forall v_j \in \mathcal{M}, \tag{11}$$

where $p = (|\mathcal{M} - 1| \operatorname{div} |\mathcal{G}|)$ and $r = (|\mathcal{M} - 1| \mod |\mathcal{G}|)$

Definition 17 (Eigenvector Centrality EV_c [3], [5]): The eigenvector centrality of a node g_i, in a graph network Υ, can be iteratively computed as:

$$EV_c(u_i) = \frac{1}{\lambda} \sum_{v_j \in \mathcal{M}} a_{u_i v_j} EV_c(v_j), \quad \forall u_i \in \mathcal{G}, \tag{12}$$

$$EV_c(v_j) = \frac{1}{\lambda} \sum_{u_i \in \mathcal{G}} a_{u_i v_j} EV_c(u_i), \quad \forall v_j \in \mathcal{M}, \tag{13}$$

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

III. BI-FACE FRAMEWORK

From a conceptual standpoint, and as depicted by the flowchart in Figure 2, the Bi-face centrality approach consists of the following basic steps.

1) We construct the formal context associated with the network and then its corresponding concept lattice. We then extract the set of bicliques that coincide with the set of formal concepts whose extent or intent is not empty.
2) We detect what we call *face-bridges*, which are the non-influential bridges in the network that contain terminal nodes.
3) We refine the bicliques by removing non-influential nodes in order to obtain *face-bicliques* (see Definition 20).
4) We compute the Bi-face centrality measures of nodes using *face-bridges* and *face-bicliques*.
5) Eventually, we use the Bi-face centrality measures to rank the two types of nodes in a descending order of importance before identifying the key ones.

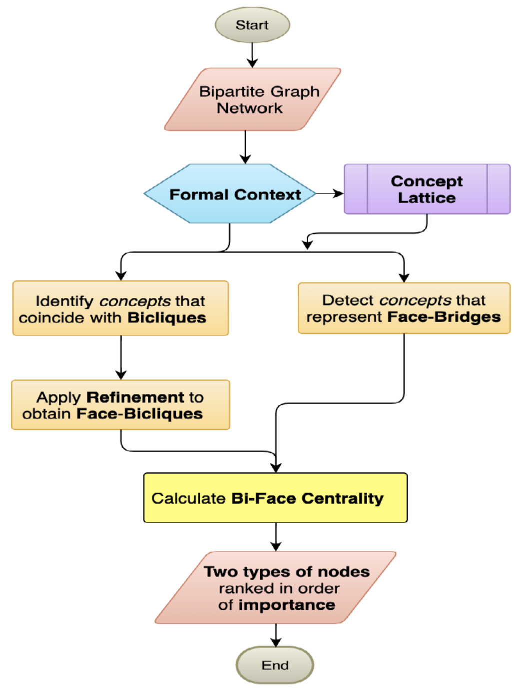


FIGURE 2. A flowchart illustrating the basic steps of the bi-face framework.

A. BUILDING THE FORMAL CONTEXT OF A TWO-MODE NETWORK

We first construct the formal context of the two-mode network $\Upsilon = (\mathcal{G}, \mathcal{M}, \mathcal{I})$ by calculating the *adjacency matrix* as follows:

$$\tilde{\mathbb{K}} = (\mathcal{G}, \mathcal{M}, \mathcal{I}) = \begin{cases} (u_i, v_j) = 1, & \exists\, (u_i, v_j) \in \mathcal{I} \\ (u_i, v_j) = 0, & \text{Otherwise.} \end{cases} \tag{14}$$

In Eq. (14), If the object u_i (node type-I) is linked to the attribute v_j (node type-II) in the network Υ, we set 1 to $\tilde{\mathbb{K}}$ element in the row i and column j. Otherwise, we assign 0 to it. For instance, Table 1 shows the constructed formal context $\tilde{\mathbb{K}}$ of our toy graph in Figure 1.

We then construct the concept lattice $\mathfrak{B}(\tilde{\mathbb{K}})$ from the formal context, as it is shown in Figure 3. Note that Figure 3 shows

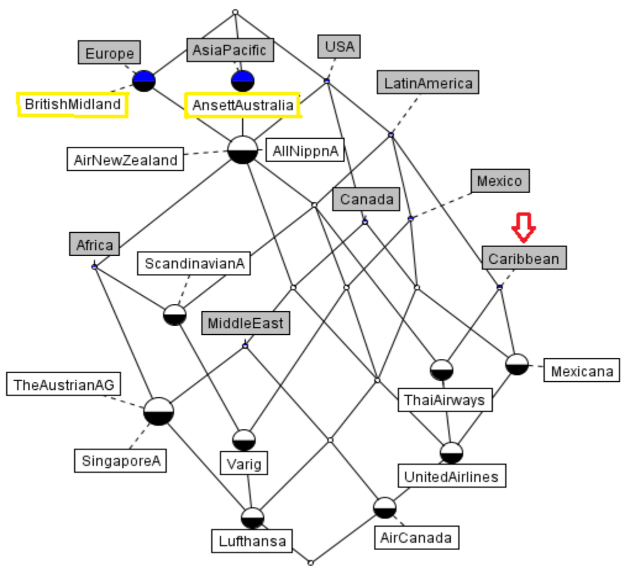


FIGURE 3. The Hasse diagram of the concept lattice $\mathfrak{B}(\tilde{\mathbb{K}})$ that corresponds to the context of the two-mode network in Figure 1. More precisely, it is a diagram with reduced labeling. This means that the label g is written below $\gamma g := (g'', g')$ and m above $\mu m := (m', m'')$. The extent of a concept represented by a node a is given by all labels in $\mathcal{G}$ from the node a downwards, and the intent by all labels in $\mathcal{M}$ from a upwards. The red downward arrow indicates the illustrative biclique cited after Definition 11.

the Hasse diagram of $\mathfrak{B}(\tilde{\mathbb{K}})$ with reduced labelling, where the label g is written below γg and m above μm. The extent of a concept represented by a node a is given by all labels in $\mathcal{G}$ from the node a downwards, and the intent by all labels in $\mathcal{M}$ from a upwards.

B. OVERLAPPING BICLIQUE EXTRACTION AND REFINEMENT

Using the constructed lattice $\mathfrak{B}(\tilde{\mathbb{K}})$, it is now possible to extract concepts that capture the corresponding bicliques of the two-mode network as follows:

Proposition 18: Given a network Υ and its corresponding concept lattice $\mathfrak{B}(\tilde{\mathbb{K}})$, a concept $c = (A, B) \in \mathfrak{B}$ with $|A| \geq 1$ and $|B| \geq 1$, represents a biclique $Q = (\{u : u \in A\}, \{v : v \in B\})$ in Υ.

Proof: Given a concept $c = (A, B) \in \mathfrak{B}(\tilde{\mathbb{K}})$, then we have from Definition 2 that $\forall u \in A, \forall v \in B, \exists (u, v) \in \mathcal{I}$. This entails that the concept c represents a sub-matrix $\hat{Q} \subseteq \tilde{\mathbb{K}}$ of size $|A| \times |B|$ that contains all 1's. Now, given that the concept lattice $\mathfrak{B}$, which is constructed from the formal context $\tilde{\mathbb{K}}$, is equivalent to the network Υ – where the sets of objects $\mathcal{G}$, attributes $\mathcal{M}$ and relations $\mathcal{I}$ in $\tilde{\mathbb{K}}$ correspond to the two disjoint sets of nodes and set of edges in Υ respectively – we can then deduce that the sub-matrix $\hat{Q}$ of the concept c coincides with a complete sub-graph $Q = (A, B)$ in Υ such that $\forall u \in A, \forall v \in B$ there is an edge (u, v) that connects the two nodes u and v. Pursuant to Definition 11, this complete sub-graph Q represents a biclique $(\{u : u \in A\}, \{v : v \in B\}) \in \Upsilon$. This implies that the concept $c = (A, B) \in \mathfrak{B}$ is equivalent to a biclique $Q \in \Upsilon$ in which both extent A and intent B involve only the objects $\{u : u \in A\}$ and attribute $\{v : v \in B\}$ nodes of Q respectively. ■

An interesting question that could be raised now is how to determine the non-influential nodes in a given concept (or biclique). To answer this question, let us define a non-influential node from the viewpoint of FCA.

Definition 19 (Non-Influential Node): For a formal concept (biclique) $c_i = (A_i, B_i) \in \mathcal{C}$, a type-I node $u \in A_i$ is *non-influential* if its removal from c_i (and accordingly from the graph $\mathcal{G}$) does not violate the closure conditions of other biclique concepts $\mathcal{C} \setminus \{c_i\}$ that involve it:

$$\forall c_j \in \mathcal{C} \setminus \{c_i\} \text{ and } u \in A_j, (A_j \setminus \{u\})'' = A_j. \tag{15}$$

In a dual manner, a type-II node $v \in B_i$ is *non-influential* if:

$$\forall c_j \in \mathcal{C} \setminus \{c_i\} \text{ and } v \in B_j, (B_j \setminus \{v\})'' = B_j. \tag{16}$$

That is, the subset of concepts (or bicliques) that contain either node u or node v nevertheless preserve their conceptual substructures even after eliminating u from their extents or v from their intents. In fact, this implies that the node u or v is non-influential (e.g., has no essential conceptual information) since removing it from the bicliques does not influence the network's intrinsic connectivity (e.g., which may clearly appear through the non-expansion of the concepts' extents or intents). In fact, Definition 19 raises another interesting question of how to identify the non-influential nodes in bicliques. Fortunately, the faces of corresponding concepts, w.r.t. their upper and lower covers, can reveal information about their non-influential nodes. As a result, one efficient way to answer this question is to juxtapose the corresponding concept (biclique) with its lower and upper covers through extensional and intentional faces to determine its non-influence type-I and type-II nodes respectively. That is, the set of faces of its concept $c_i = (A_i, B_i)$, w.r.t. its lower and upper covers, have in common the same non-influential (type-I and type-II) nodes in its (extent and intent) respectively:

$$\begin{aligned} \forall u \in \{\cap_{c_l \in \mathfrak{B}(c_i)} f_{\text{ex}}(c_i, c_l)\} &\implies (A_j \setminus \{u\})'' = A_j, \\ &\forall c_j \in \mathcal{C} \setminus \{c_i\} \text{ and } u \in A_j. \end{aligned} \tag{17}$$

$$\begin{aligned} \forall v \in \{\cap_{c_d \in \mathcal{U}(c_i)} f_{\text{in}}(c_i, c_d)\} &\implies (B_j \setminus \{v\})'' = B_j, \\ &\forall c_j \in \mathcal{C} \setminus \{c_i\} \text{ and } v \in B_j. \end{aligned} \tag{18}$$

For example, the corresponding concept of $\tilde{Q}$ has two extensional faces $f_{ex}^1 = \{ThaiAirways\}$ and $f_{ex}^2 = \{Mexicana\}$. Since the intersection of the faces f_{ex}^1 and f_{ex}^2 is empty, $\tilde{Q}$ contains no non-influential type-I nodes. It also has only one intensional face $f_{in}^1 = \{Caribbean\}$. Thus, the intersection is also f_{in}^1, which entails that *Caribbean* is a non-influential type-II node in the $\tilde{Q}$.

On the basis of Equations (17) and (18), we can leverage the faces of concepts to define a key biclique[2] as follows:

Definition 20 (Face Biclique): Given a two-mode network Υ and its corresponding concept lattice $\mathfrak{B}(\tilde{\mathbb{K}})$, a concept

[2]Note that a biclique is key when all of its nodes are influential.

(representing a biclique) $c = (A, B) \in \mathfrak{B}$, is called a *face biclique* if all of its (type-I and II) nodes are influential, i.e., no one of them satisfies the conditions in Equations (17) and (18).

Based on Definition 20, we can obtain the face biclique $\hat{c} = (\hat{A}, \hat{B})$ by refining the original biclique $c = (A, B)$ as follows:

$$\hat{A} = \begin{cases} A \setminus \{\cap_{c_l \in \mathcal{L}(c_i)} f_{\text{ex}}(c_i, c_l)\}, & |A| > 1 \\ A, & \text{otherwise,} \end{cases}$$
$$\hat{B} = \begin{cases} B \setminus \{\cap_{c_d \in \mathcal{U}(c_i)} f_{\text{in}}(c_i, c_d)\}, & |B| > 1 \\ B, & \text{otherwise.} \end{cases} \quad (19)$$

In Equation (19), we remove non-influential type-I nodes from its extent and non-influential type-II nodes from its intent. It is worth noting that when the extent or intent contains only one node, no refinement is applied because this node is influential by default. This is due to the fact that removing this node clearly violates the closure conditions in Equations (17) and (18).

C. FACE-BRIDGE DETECTION

Definition 21 (Face-I Bridge and Terminal Type-I Node): Given a 2-mode network Υ and its corresponding concept lattice $\mathfrak{B}(\tilde{\mathbb{K}})$, an edge (u, B) represents a non-influential (face-I) bridge containing a terminal (type-I) node $u \in \mathcal{G}$ when there is an attribute concept $c = (A, B) \in \mathfrak{B}(\tilde{\mathbb{K}})$ with $|B| = 1$ that satisfies the following:

$$u \in A \ \text{ and } \ \exists h_i \in \mathcal{H}_c^{\text{ex}} \quad \text{S.t. } h_i = u \text{ and } |h_i| = 1 \quad (20)$$

For instance, the attribute concept $c =$ ({AirCanada, AirNewZealand, AllNippnA, TheAustrianAG, BritishMidland, Lufthansa, ScandinavianA, SingaporeA, ThaiAirways, UnitedAirlines, Varig}, {Europe}) that appears in blue/black in Figure 3 has an extensional minimal generator set $\mathcal{H}_c^{\text{ex}}$ = {BritishMidland}. This implies that *BritishMidland* (framed in yellow in Figure 3) is a terminal (type-I) node and the edge (BritishMidland, Europe) represents a non-influential (face-I) bridge. Similarly, we have:

Definition 22 (Face-II Bridge and Terminal Type-II Node): Given a 2-mode network Υ and its corresponding concept lattice $\mathfrak{B}(\tilde{\mathbb{K}})$, an edge (A, v) represents a non-influential (face-II) bridge containing a terminal type-II node $v \in \mathcal{M}$ when there is an object concept $c = (A, B) \in \mathfrak{B}(\tilde{\mathbb{K}})$ with $|A| = 1$ that satisfies the following:

$$v \in B \ \text{ and } \ \exists h_j \in \mathcal{H}_c^{\text{in}} \quad \text{S.t. } h_j = v \text{ and } |h_j| = 1 \quad (21)$$

The question now is, how can we obtain the minimal generators of object and attribute concepts? We can efficiently compute the set of minimal generators $\mathcal{H}_c^{\text{in}}$ of a concept c intent by applying *Minigen()* procedure, which is given in Algorithm 1. It iteratively calculates the face of c w.r.t. each upper cover in $\mathcal{U}(c)$ (Line 3). If the set of intentional minimal generators is empty, it then assigns the individual attributes in the first face to $\mathcal{H}_c$ (Lines 4-5). Otherwise, it progressively checks the intersection between the calculated face f_u and

Algorithm 1 *Minigen()* Procedure for Computing the Intentional Minimal Generators of a Concept Intent

Input: Concept intent B, Set of upper covers $\mathcal{U}(c)$.
Output: Set of minimal generators $\mathcal{H}_c^{\text{in}}$.

1: $\mathcal{H}_c^{\text{in}} \leftarrow \emptyset$;
2: **for** each $c_u = (A_u, B_u)$ in $\mathcal{U}(c)$ **do**
3:   $f_u \leftarrow B \setminus B_u$;
4:   **if** $\mathcal{H}_c^{\text{in}} == \emptyset$ **then**
5:     $\mathcal{H}_c^{\text{in}} \leftarrow \{a | \forall a \in f_u\}$;
6:   **else**
7:     Gen $\leftarrow \emptyset$;
8:     **for** each h_i in $\mathcal{H}_c^{\text{in}}$ **do**
9:       **if** $h_i \cap f_u == \emptyset$ **then**
10:         Gen $\leftarrow$ (Gen $\cup \{h_i \cup a | \forall a \in f_u\}$);
11:       **else**
12:         Gen $\leftarrow$ (Gen $\cup \{h_i\}$);
13:       **end if**
14:     **end for**
15:     $\mathcal{H}_c^{\text{in}} \leftarrow$ minimal(Gen);
16:   **end if**
17: **end for**
18: **Return** $\mathcal{H}_c^{\text{in}}$;

each generator h_i in $\mathcal{H}_c^{\text{in}}$ (Line 8). If the intersection with the current generator h_i is empty, then h_i is not in the family blocker formed by the face (Line 9). This entails that the generator h_i must then be modified so that it belongs to the minimal blocker family of faces. Thus, the new minimal generators will be obtained by adding each element of the current face f_u to h_i (Line 10). If the intersection is not empty, then the current generator h_i, which exists in the family of minimal blockers of previous faces, is also a minimal blocker of the family formed of the current face f_u. So, we add the generator h_i, without performing any modification to the minimal generator set $\mathcal{H}_c^{\text{in}}$ (Line 12). It ultimately verifies the minimality of the obtained set (Line 15) and returns the final set of minimal generators $\mathcal{H}_c^{\text{in}}$ (Line 18). Note that, in a dual way and using the set of concept's lower-covers $\mathcal{L}(c)$, we can apply *Minigen*() procedure to compute the set of extensional minimal generators $\mathcal{H}_c^{\text{ex}}$ of a concept w.r.t. its extent A.

D. BI-FACE CENTRALITY

Definition 23 (Bi-face Centrality BF_c): The Bi-face centrality of nodes $u \in \mathcal{G}$ and of $v \in \mathcal{M}$, in a given graph network Υ, can be computed as:

$$\text{BF}_{\text{I}}(u) = \frac{\overbrace{|\{\hat{c} \in \hat{\mathcal{C}} \mid u \in \hat{A}\}|}^{\text{Face-bicliques containing } u}}{|\hat{\mathcal{C}}|} + \Big[1 - \frac{\overbrace{|\{g \in \Gamma_I \mid g == u\}|}^{\text{Face-I bridges containing } u}}{|\Gamma_I|}\Big], \quad (22)$$

$$\mathrm{BF}_{\mathrm{II}}(v) = \frac{\overbrace{|\{\hat{c} \in \hat{\mathcal{C}} \mid v \in \hat{B}|\}}^{\text{Face-bicliques containing } v}}{|\hat{\mathcal{C}}|} + \big[1 - \frac{\overbrace{|\{m \in \Gamma_{II} \mid m == v|\}}^{\text{Face-II bridges containing } v}}{|\Gamma_{II}|}\big]. \quad (23)$$

$\hat{\mathcal{C}}$ stands for the set of face bicliques while Γ_I and Γ_{II} represent the two sets of non-influential (face-I) and (face-II) bridges, respectively. In Eq. 22, the Bi-face centrality calculates the sum of *face-biclique*[3] and *face-bridge* terms. The numerator of the face-biclique term simply counts the number of refined concepts, with extent and intent sizes greater than 1, that involve a type-I node u. That is, it measures the amount of face bicliques to which node u belongs in the graph network Υ. From a conceptual viewpoint, this term effectively approximates the cross connectivity [20], [21] of the node u using refined overlapped bicliques that only contain influential nodes. In the face-bridge term, we first quantify the ratio of the face bridges that involve the node u. This ratio is then subtracted from 1 to approximate the portion of influential bridges in the graph that contain the node u. It is worth noting that the numerators of two Bi-Face terms in Eq. (22) are unnormalized quantities. As a result, the denominators in Eq. 22 act as normalization factors to scale the two terms between 0 and 1. In a similar manner, the Bi-face centrality in Eq. (23) can be interpreted and used to compute the centrality of type-II nodes in the graph.

The pseudo-code for calculating the Bi-face centrality of all type-I nodes in the two-mode network Υ is given in Algorithm 2. The algorithm takes as input the set of all extracted concepts $\mathcal{C} = \{c_j = (A_j, B_j)\}_{j=1}^{|\mathcal{C}|}$. For each type-I node $u_i \in \mathcal{G}$, it first iteratively refines the extents of the bicliques to obtain the face ones by removing all their non-influential type-I nodes (lines 4-5). It then counts the number of those refined face bicliques in the graph that involve u_i (lines 7-9). Hereafter, it iteratively computes the minimal generators of the attribute concepts w.r.t. their extents to identify the face-bridges that involve the node u_i (lines 11-12). Subsequently, it counts how many face-bridges containing the node u_i as a terminal (type-I) one (lines 13-15). It then calculates the Bi-face centrality BF_I of a node u_i (lines 19-21). Eventually, it returns a list with the Bi-face centrality measures BF_I of all type-I nodes in the graph respectively (line 22). Without loss of generality, and in a dual manner, algorithm 2 can be applied to compute the Bi-face centrality for each type-II node $v_j \in \mathcal{M}$ as follows. It iteratively obtains the face bicliques by refining the non-influential type-II nodes from the intents of their corresponding concepts. It then identifies the face bicliques in the graph that involve v_j. It then uses the minimal generators of object concepts to count the number of the face-bridges that involve the node v_j as a terminal (type-II) one. Eventually, it returns a list containing the Bi-face centrality measures $\mathrm{BF}_{\mathrm{II}}$ of all type-II nodes in the graph.

[3]Note that the face-clique of a node is the number of overlapping face bicliques to which it belongs to.

Consequently, we can now use the resulting Bi-face centrality lists to rank the two types of nodes in descending order based on their importance. Table 2 summarizes the ranked lists of the most important airlines and destinations, in Figure 1, based on five bipartite centrality measures: Bi-face, betweenness, eigenvector, closeness and degree. For example, because the node *Lufthansa* has slightly fewer geodesics than *AirCanada*, Betweenness considers *AirCanada* to be the most important type-I node. In contrast, the Bi-face centrality ranks the node *Lufthansa* as the most important type-I node because *Lufthansa* exists in considerably more overlapped bicliques than *AirCanada*. Closeness, degree, and eigenvectors are unable to distinguish which node Lufthansa or *AirCanada* is more important than another. Furthermore, neither degree nor closeness centrality can determine which type-I node from {*TheAustrianAG*, *SingaporeA*, *Varig*} is more influential than the others. The eigenvector centrality cannot distinguish between type-II nodes in {*MiddleEast*, *Africa*, *Caribbean*}.

Algorithm 2 Calculating Bi-Face Centrality ($\mathrm{BF_c}$) for All Type-I Nodes in a Two-Mode Network

Input: Set of bicliques ($\mathcal{C} = \{(A_j, B_j)\}_{j=1}^{|\mathcal{C}|}$).
Output: Bi-face centrality ($\mathrm{BF_I}$) of all type-I nodes.

1: $\mathrm{BF_I} \leftarrow \Gamma_I \leftarrow \emptyset$;
2: **for** each $u_i \in \mathcal{G}$ **do**
3:   $\mathrm{count}_I \leftarrow \gamma_I \leftarrow [0]_{i=1}^{|\mathcal{G}|}$;
4:   **for** each $A_j \in \mathcal{C}$ **do**
5:     $\hat{A}_j \leftarrow \mathrm{Refine}(A_j)$; //*usingEq*. 19
6:     *// Counting face bicliques that contain the node* u_i
7:     **if** $|\hat{A}_j| > 0$ and $u_i \in \hat{A}_j$ **then**
8:       $\mathrm{count}_I[i] \leftarrow \mathrm{count}_I[i] + 1$;
9:     **end if**
    // Counting face-bridges that contain the node u_i
10:     **if** $|B_j| == 1$ **then**
      // using the extensional version of Algorithm 1.
11:       $\mathcal{H}_{A_j}^{\mathrm{ex}} \leftarrow Minigen(A_j)$;
12:       **if** $\exists h \in \mathcal{H}_{c_j}^{\mathrm{ex}}, h == u_i$ **then**
13:         $\gamma_I[i] \leftarrow \gamma_I[i] + 1$; $\Gamma_I.append(u_i)$;
14:       **end if**
15:     **end if**
16:   **end for**
17: **end for**
18: **for** each $i = 1$ to $|\mathcal{G}|$ **do**
19:   $\mathrm{BF_I}[i] \leftarrow \big(\mathrm{count}_I[i]/|\mathcal{C}|\big) + \big(1 - (\gamma_I[i]/|\Gamma_I|)\big)$;
20: **end for**
21: **Return** $\mathrm{BF_I}$

1) COMPLEXITY ANALYSIS

The calculation of the face biclique term has a time and a space complexity equal to $O(|\mathcal{C}|)$ since we store and proceed through the extent of all the bicliques to count

TABLE 2. The ranking of all nodes in the two-mode network of Figure 1 based on five bipartite centrality measures: Bi-face (BF_c), betweenness (B_c), eigenvector (E_c), closeness (C_c) and degree (D_c).

	Rank	**Type-I**	**Type-II**
BFc	**1**	Lufthansa	USA
	2	AirCanada	AsiaPacific, Europe
	3	UnitedAirlines	LatinAmerica
	4	Varig	Canada
	5	SingaporeA, TheAustrianAG	Mexico
	6	ScandinavianA	Africa
	7	ThaiAirways	MiddleEast
	8	Mexicana	Caribbean
	9	AirNewZealand, AllNippnA	-
	10	AnsettAustralia, BritishMidland	-
Ec	**1**	Lufthansa, AirCanada	USA
	2	UnitedAirlines	AsiaPacific, Europe
	3	Varig	Canada
	4	SingaporeA, TheAustrianAG	LatinAmerica
	5	ThaiAirways, ScandinavianA	Mexico
	6	Mexicana	MiddleEast, Africa, Caribbean
	7	AllNippnA, AirNewZealand	-
	8	AnsettAustralia, BritishMidland	-
Cc	**1**	Lufthansa, AirCanada	USA, AsiaPacific, Europe
	2	UnitedAirlines	LatinAmerica
	3	Varig, SingaporeA, TheAustrianAG	Canada
	4	ThaiAirways, ScandinavianA	Mexico
	5	Mexicana, AllNippnA, AirNewZealand	MiddleEast, Africa
	6	AnsettAustralia, BritishMidland	Caribbean
Bc	**1**	AirCanada	AsiaPacific, Europe
	2	Lufthansa	USA
	3	UnitedAirlines	LatinAmerica
	4	SingaporeA, TheAustrianAG	Canada
	5	Varig	Mexico
	6	ThaiAirways	Africa
	7	ScandinavianA	Caribbean
	8	Mexicana	MiddleEast
	9	AirNewZealand, AllNippnA	-
	10	AnsettAustralia, BritishMidland	-
Dc	**1**	AirCanada, Lufthansa	Europe, AsiaPacific, USA
	2	UnitedAirlines	LatinAmerica
	3	TheAustrianAG, SingaporeA, Varig	Canada
	4	Mexicana, ScandinavianA, ThaiAirways	Africa, Mexico
	5	AirNewZealand, AllNippnA	MiddleEast, Caribbean
	6	AnsettAustralia, BritishMidland	-

the face bicliques that contain the node. The Face-bridge term of type-I node needs iterating through the attribute concepts $\tilde{\mathcal{C}}$ and calculates their minimal generators w.r.t. their corresponding lower covers. Thus, the Bi-face centrality BF_I of all type-I nodes requires $\big(|\mathcal{G}| \times |\mathcal{C}| + |\tilde{\mathcal{C}}| \times |\tilde{\mathcal{L}}| \times |\tilde{\mathcal{H}}^{ex}|\big)$, where $\tilde{\mathcal{C}}$ is the set of attribute concepts, $|\tilde{\mathcal{H}}^{ex}|$ is the largest size of an obtained set of minimal generators for attribute concepts, and $\tilde{\mathcal{L}}$ is the largest number of lower covers for an attribute concept. Now, since we often have $|\tilde{\mathcal{C}}| \ll |\mathcal{C}|$ and also $|\tilde{\mathcal{L}}| \ll |\mathcal{G}|$, then the first term frequently dominates the second one. This entails that computing the Bi-face centrality BF_I of all type-I nodes needs a time and space complexity of $O(|\mathcal{G}| \times |\mathcal{C}|)$. In a dual way, the calculation of the Bi-face centrality BF_{II} of all type-II nodes has a time complexity of $O(|\mathcal{M}| \times |\mathcal{C}|)$. In total, the Bi-face centrality has time and space complexity of $O\big(|\mathcal{C}| \times (|\mathcal{G}| + |\mathcal{M}|)\big)$.

2) BI-FACE VS. CROSS-FACE

One might contrast the bipartite Bi-Face (BF) and our Cross-face (CF) centrality introduced in [22], which is a prominent FCA-based centrality for one-mode networks. At a high level, BF can be considered as a generalized form of CF (with a larger size and a higher level of details and depth) for two-mode networks. However, it is well-known that two-mode networks have distinct characteristics with more complex substructures than the one-mode ones, which leads to a different computation of node centrality and distinct applications. Thus, recalling the basic formulations used in both centrality approaches, the BF and CF are fundamentally different measures that share a similar FCA-based route. Technically, some of these differences can be summarized as follows: (1) In the preprocessing step of BF, a different adjacency matrix (see Eq. (14)) adapted for two-mode networks is used to build the formal context; (2) In the BF framework, we extract bicliques and bridges using concepts rather than symmetrical concepts as in the CF framework. This is due to the fact that the symmetrical ones do not exist in the constructed lattice representing two-mode networks; (3) In the BF approach, we use a refinement step to obtain what we name face bicliques by pruning the non-influential nodes from the original bicliques (see Definition 20 and Eq. (19)). This step does not exist in the CF one-mode formulation centrality. In practice, this step significantly improves the outcome (as demonstrated by the results of experiment I in subsection IV-C); (3) Face bridges in BF are identified in a completely different way than in one-mode networks for CF. Specifically, we leverage the minimal generators of concepts in BF (see Definitions 21-22 and Eqs. (20)-(21)) instead of the interesting faces and meet-irreducible concepts in CF; (4) We detect the terminal nodes in the BF framework, which is not taken into account in CF. Due to the aforementioned differences, the two terms (i.e., face-bicliques and face-bridges in Eqs. (20)-(21)) formulations of BF meaningfully differ from the corresponding terms of CF formulation presented in [22].

IV. EXPERIMENTAL EVALUATION

The objective of our experimental evaluation is to find robust answers to the following essential questions.

- (**Q1**) Is the accuracy of Bi-face centrality competitive with the prominent centrality measures?
- (**Q2**) Is Bi-face centrality performing fast compared to state-of-the-art centrality measures?
- (**Q3**) Is there a correlation between the Bi-face centrality approach and other state-of-the-art centrality measures?

We first consider the following five (real-life* and synthetic‡) two-mode networks, which possess various configurations to support the investigation of different scenarios.

A. DATASETS

- ***Norwegian Interlocking Directorates** [23], which contains interlocking boards of 1542 Norwegian director women in 373 Norwegian public limited companies. A link represents a board membership connecting a woman as a director of a public company in Norway on August 2009.
- ***PediaLanguages** [24] involves the semantic web of 316 official languages spoken by people living in 169 different countries. An edge connects an official language to a country if people in that country speak that language.
- ***Southern-Women-Davis** [25], [26], which is a two-mode social network of 18 women reporting their participation to 14 events (such as a meeting of a social club, a church event or a party) over a nine-month period. A woman is connected to an event if she attends that event.
- ‡**CoinToss**, which is a random bipartite network generated from indirect Coin-Toss model generator [27].
- ‡**Dirichlet** [28] which is a random formal context generated using the *Dirichlet* model generator.[4]

Table 3 gives the basic statistics of the networks.[5]

[4]publicly available at: https://github.com/maximilian-felde/formal-context-generator

[5]Publicly available at: https://toreopsahl.com/datasets/
http://konect.cc/networks/opsahl-collaboration/
https://networkdata.ics.uci.edu/netdata/html/davis.html

TABLE 3. The basic statistics of the two-mode networks about the number $|\mathcal{G}|$ of type-I nodes, the number $|\mathcal{M}|$ of type-II nodes, the number $|\mathcal{I}|$ of edges, and the density Θ in %.

Name	$\|\mathcal{G}\|$	$\|\mathcal{M}\|$	$\|\mathcal{I}\|$	Θ
Norwegian-Directorate	1542	375	1889	0.33
PediaLanguages	316	169	9022	17
Southern-Women-Davis	18	14	89	36
CoinToss	793	10	3310	42
Dirichlet	1100	55	30855	51

B. METHODOLOGY

The results of our proposed **Bi-face** centrality measure are then compared to the following state-of-the-art measures:

- **Bipartite closeness** [Definition 15].
- **Bipartite betweenness** [Definition 16].
- **Bipartite eigenvector**[Definition 14].
- **Vote-Rank** [29], which is a well-known method for identifying decentralized spreaders. It calculates the ranking of the nodes in the bipartite graph based on a voting scheme. That is, at each turn, all nodes iteratively vote in a spreader. The node with the highest vote number is elected iteratively, while decreasing the voting ability of the elected spreader' neighbours in the next turn.
- **Percolation** [11], which measures the proportion of percolated paths[6] that go through a given node. So, it quantifies the relative impact of nodes in various percolation scenarios based on their topological connectivity over time. The percolation state is commonly assigned a value between 0.0 and 1.0, with 0.5 being the most commonly used value that we used in our experiment.
- **Bipartite degree** [Definition 14], which can act as an effective baseline for comparison.

Subsequently, the two ranking lists of (type-I and type-II) nodes calculated from the underlying centrality measures are then compared with the corresponding lists obtained from the spreading process of the node. We specifically evaluate the tested centrality's performance for each type of node by applying the following common schema [30], [31]:

1) Calculate the centrality metric for all nodes and record their ranking list
2) Simulate the spreading ability of nodes using SIR model [30]. The node in the SIR model can be susceptible, infected, or recovered. We set only one node to be infected at a time, and the other remaining nodes are susceptible, then we examine how the information spreads on the network. With a spreading (or infection) probability, the infected node can spread its infection to nearby susceptible nodes. In practice, we noticed that investigating the spreading in the early stages is more meaningful than examining each node recovered state, so we concentrate on the effect within a $t = 10$ time range rather than the recovered state of each

[6]We recall that the percolated path is the shortest one between two nodes in which the source node is percolated (i.e., infected).

node. Following the completion of the SIR simulation, we obtain the node influence ranking list by computing the spreading efficiency for all nodes.

3) Compute the joint score list $\mathcal{J} = \{(x_i, y_i)\}_{i=1}^{n}$ using the SIR model's ranking list and the centrality measure's ranking list. The x_i and y_i in each pair $(x_i, y_i) \in \mathcal{J}$ are the centrality-based and SIR-based measures of a node $g_i \in \mathcal{G}$, respectively. The two randomly chosen pairs $(x_i, y_i), (x_j, y_j) \in \mathcal{J}$ are *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 *discordant* if both $(x_i < x_j)$ and $(y_i > y_j)$ or if both $(x_i > x_j)$ and $(y_i < y_j)$. If $(x_i = x_j)$ and $(y_i = y_j)$, then the pair is neither concordant nor discordant. We use n_c and n_d to denote the number of concordant and discordant pairs in $\mathcal{J}$, respectively.

Based on $\mathcal{J}$, we then calculate the following Kendall's tau rank correlation coefficient τ metric:

$$\tau = \frac{2(n_c - n_d)}{n(n-1)}, \tag{24}$$

If the underlying centrality measure has a high τ value, this indicates that it produces an accurate ranked list. The ranked list produced by the centrality measure is identical to the ranked list obtained from the real spreading process when $\tau = 1$, which is, in fact, the ideal scenario. To evaluate the accuracy of the results, we now calculate the average Kendall's tau rank correlation coefficient as follows:

$$\hat{\tau} = \frac{\tau_I + \tau_{II}}{2}, \tag{25}$$

where τ_I and τ_{II} are the Kendall's tau correlation coefficients calculated using Eq. (24) for type-I and type-II of nodes, respectively.

To assess the scalability, we consider the average elapsed time metric as:

$$\xi = \frac{1}{2}\left[\frac{\sum_{u_i \in \mathcal{G}} t_i}{n} + \frac{\sum_{v_j \in \mathcal{M}} t_j}{m}\right] \tag{26}$$

where t_i and t_j are the elapsed times for calculating the centrality measure of a type-I node $u_i \in \mathcal{G}$ and a type-II one $v_j \in \mathcal{M}$, respectively.

We carried out our experiments on a MacOS Mojave computer with an Intel(R) Core-i7 CPU @2.6GHz and 16 GB of memory. As an extension to the NetworkX Python package, we implemented all of the centrality measures. We also used the *Concepts 0.7.11* Python package, developed by Sebastian Bank,[7] to extract formal concepts.

C. RESULTS

1) EXPERIMENT I

This experiment is devoted to answering Question 1. Each infected node has a spreading probability β of infecting its susceptible neighbours in the SIR model simulation. As a result, and in accordance with the scheme described above, we iteratively increase the spreading probability in the range $\beta = (0, 0.1]$ with increments of 0.01. At each step-size, we compute the joint list $\mathcal{J}$ of each centrality measure and the real spreading of the nodes for each individual type of nodes separately. We then calculate the corresponding evaluation metric $\hat{\tau}$ in Eq. (25).

[7] publicly available at: https://pypi.python.org/pypi/concepts

Figure 4 displays the average Kendall's tau correlation coefficient $\hat{\tau}$ between 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, Bi-face outperforms all the compared centrality measures, achieving the most accurate Kendall coefficient $\hat{\tau}$ on *Norwegian-Directorate*, *PediaLanguages*, *CoinToss* and *Dirichlet* networks. On the *Women-Davis* network, Bi-face has the highest $\hat{\tau}$ value when the spreading probability $\beta \geq 0.03$, otherwise vote-rank, closeness, betweenness and degree slightly compete with Bi-face. The percolation comes close behind Bi-face on *Women-Davis*, but considerably further behind on *Norwegian-Directorate*, *PediaLanguages*, *CoinToss* and *Dirichlet* networks. Except on the *Women-Davis* network with spreading probability $\beta < 0.03$, the vote-rank is clearly less accurate than Bi-face on all the tested networks, but it is more accurate than percolation, betweenness, closeness, eigenvector and degree on *PediaLanguages*, *CoinToss* and *Dirichlet* networks. On the *Norwegian-Directorate* and *Women-Davis* networks, the vote-rank and percolation compete with each other. The percolation is clearly more accurate than betweenness and eigenvector when the spreading probability $\beta \geq 0.05$ on all the tested networks. Both betweenness and eigenvector dominate degree and closeness on *Norwegian-Directorate*, *PediaLanguages* and *CoinToss* networks, but closeness outperforms betweenness centrality on *Dirichlet* network. The betweenness is more accurate than eigenvector on *PediaLanguages* network when the spreading probability $\beta \geq 0.04$, but it is outperformed by eigenvector on *CoinToss* and *Dirichlet* networks. The betweenness centrality is almost better than eigenvector for the *Norwegian-Directorate* network while they have an opposite behavior for the *Women-Davis* network.

2) EXPERIMENT II

The second experiment is dedicated to answer Question 2. The goal here is to evaluate the performance of the centrality measures. To that end, we rerun Experiment I while reporting their computational time as in Eq. 26. The average elapsed time ξ of the seven centrality measures on the five underlying networks is depicted in Figure 5. On all the tested networks, the Bi-face dominates all centrality measures (except degree). It finishes at least twenty-three times faster than betweenness, eleven times faster than percolation, nine times faster than eigenvector and ten times faster than closeness. Degree is very competitive with Bi-face on *Women-Davis*, *CoinToss* and *Dirichlet* networks, but Bi-face clearly prevailed over the degree by a significant margin on *Norwegian-Directorate* and *PediaLanguages* networks. Apart from

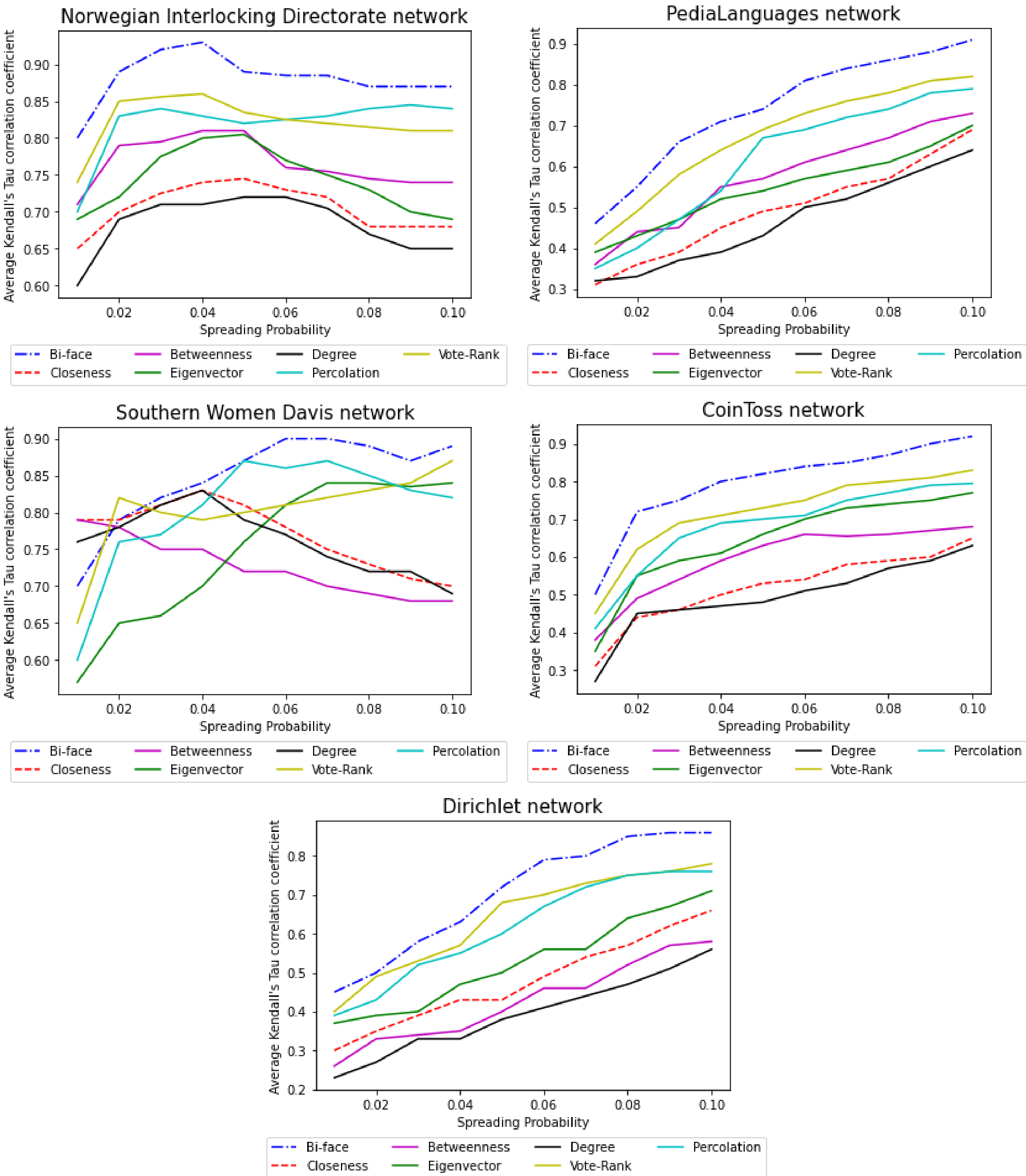


FIGURE 4. The average Kendall's tau coefficient $\hat{\tau}$ between the tested centrality measures and the ranking list generated by the SIR model, with $\beta \in (0, 0.1]$, at $t = 10$ on the five underlying datasets.

Bi-face, the percolation is marginally slower than both the closeness and eigenvector by at least factors of 1.15 and 1.25 on all networks (except *Women-Davis*) respectively. In addition, the closeness is considerably faster than betweenness, and competes with eigenvector on *Norwegian-Directorate* and *CoinToss* networks. Vote-rank is significantly faster than closeness, eigenvector and percolation on *Norwegian-Directorate*, *PediaLanguages*, *CoinToss* and *Dirichlet* networks, but on the contrary, closeness is slightly quicker than vote-rank on *Women-Davis* network.

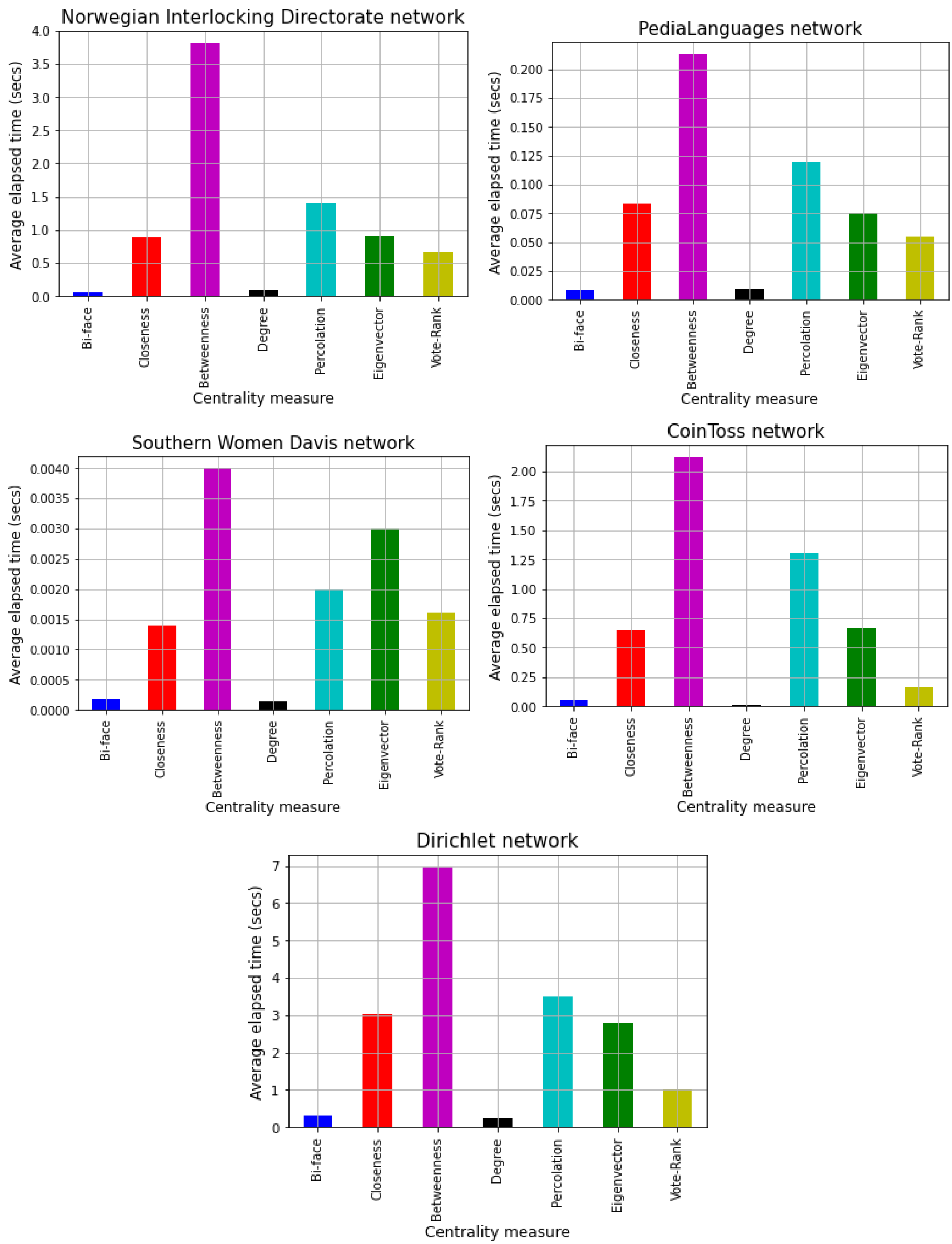


FIGURE 5. Average elapsed time ξ (in secs) of the seven tested centrality measures: Bi-face, closeness, betweenness, degree, eigenvector, percolation and vote-rank on the five underlying datasets.

3) EXPERIMENT III

In this experiment, we focus on Question 3. That is, we are interested here in exploring the monotonic relationships between Bi-face and the other underlying centrality measures. Table 4 records the average Kendall's tau rank correlation coefficient between Bi-face and the other six bipartite centrality measures. Overall, all the centrality measures are positively correlated with Bi-face, which is remarkably consistent and supplement the finding of Experiment I. The Bi-face has moderate monotonic relationships with vote-rank and percolation on all tested networks. There is clearly a weak relationship between Bi-face and betweenness on the *Norwegian-Directorate*, *PediaLanguages*, and CoinToss networks. Furthermore, it has a weak correlation with eigenvector on the *Women-Davis*, *CoinToss* and *Dirichlet* networks. Moreover, there is weak correlation between Bi-face and closeness on the *Women-Davis* and *Dirichlet* networks. Noticeably, the Bi-face has a very weak relationship with degree on all networks except the *Women-Davis*.

D. DISCUSSION

Taking the identification of accurate node centrality into consideration, the results of Experiment I in Subsection IV-C1 indicate that Bi-face outperforms traditional bipartite centrality measures such as vote-rank, percolation, degree, closeness, betweenness, and eigenvector. This is attributed to the use of its face biclique and face-bridge terms in tandem to leverage local and global aspects of network topology, respectively. That is, the face-biclique term quantifies the structural embeddedness of cohesive regions in a network involving each individual (type-I and type-II) node. From a conceptual perspective, this term considers the local information on how the node influences its immediate important neighbour nodes through the lens of its overlapping face bicliques. The face-bridge term quantifies a node's global role based on how the information flows through influential (face) bridges (i.e., important geodesics).

In terms of effective performance, the results of Experiment II from the previous Subsection IV-C2, suggest that the Bi-face is considerably faster than all other tested bipartite centrality measures (except degree). This is due to the fact that Bi-face primarily calculates the centrality of all nodes based on the set of concepts $\mathcal{C}$, which is frequently too small in comparison to all other tested centrality measures with polynomial time complexity in terms of nodes and edges, i.e., $|\mathcal{C}| \ll n^p$ and $|\mathcal{C}| \ll m^q$, with $p, q > 1$.

Besides that, several well-known observations are clearly consistent with the obtained results in Subsection IV-C. First, in some real-world applications, we may end up with several nodes having approximately similar low or high degrees, and in these cases, degree centrality cannot serve as a descriptive measure that can distinguish between nodes. Second, closeness can address the degree centrality limitation in a few situations. For example, consider node u that is linked to node v. Assume that node v is in close proximity to the other nodes in the network, resulting in a high closeness score. Node u has a very low degree score of 1, but a rationally high closeness score, because node u can propagate information to all other nodes that node v reaches with one extra step. However, closeness, like degree, is usually inappropriate for irregularly connected bipartite networks. Because the shortest-path distance between two nodes is infinite when they are not reachable through a path, the closeness score is equal (or very close) to zero for those nodes in the network that do not reach all other nodes. Third, since betweenness lacks any form of measuring local nodal connectivity, it is expected to produce relevant results only if the goal is only to quantify influence on communication among local groups, which is not always the case when studying the centrality in real-world networks. Finally, and in practice, using the efficient implementation adopted from the fastest algorithm proposed in [7], the calculation of percolation centrality for all nodes requires a time complexity of $O(m^2(n_1+n_2))$, which still seems to impose a computational bottleneck even with fairly medium-sized networks.

As frequently asked, are these centrality measures correlated? The results of Experiment III in Subsection IV-C3 expound that Bi-face centrality gives unique node identification based on network topology. The presence of terminal nodes, influential (also known as face) bridges, and overlapping key bicliques impacts both the performance and behaviour of Bi-face as well as its relationship to other traditional centrality measures. When the network contains a large number of cohesive regions with many nodes having high degrees and there is a small number of hole structures or terminal nodes, the role of the face-biclique term dominates the face-bridge one, and here it is anticipated that the Bi-face centrality could be partially correlated with vote-rank, degree, eigenvector and (maybe) closeness centrality measures. This is due to the fact that in this scenario, the network tends to decompose into multiple bi-clusters (or two-mode communities), with the nodes with the highest degree potentially serving as the central nodes. On the flip side of the coin, when the network contains a small number of cohesive regions or a large number of sparse ones, as well as a large number of terminal nodes and bridges, the role of the face-bridge term dominates the face-biclique one, even when structural holes are present. This is due to the effect of face-bridges in determining the central nodes, and here the Bi-face centrality may be slightly correlated with percolation and betweenness.

It is worth noting that the existence of the two scenarios, mentioned above in the network, could potentially increase Bi-face centrality to behave slightly similar to vote-rank or percolation. In an extreme scenario, such as the *Women-Davis* network with a large number of overlapping bicliques and no terminal nodes, the likelihood of having face-bridges decreases dramatically. This indeed imposes a harsh situation on Bi-face because it will depend solely on its face biclique term, and here it is clearly expected that Bi-face will behave similarly to degree, closeness, eigenvector, and vote-rank, but not similarly to betweenness. From a statistical perspective, the low and moderate (i.e., not high) correlations between Bi-face and other centrality measures suggest that it is, in fact, a distinct measure that is likely to be associated with different outcomes than other centrality measures. This is due to the fact that if the measures are highly correlated, they may be somewhat redundant and behave similarly.

TABLE 4. Average Kendall's tau rank correlation coefficient between Bi-face (BF_c) and the other six bipartite centrality measures: betweenness (B_c), eigenvector (E_c), closeness (C_c), degree (D_c), Percolation (PC_c) and Vote-rank (VR_c) on the five underlying datasets. The moderate, weak, highly weak correlation values are represented in blue, red and black respectively.

	BF_c				
	Norwegian-Directorate	PediaLanguages	Women-Davis	CoinToss	Dirichlet
E_c	0.09	0.12	0.19	0.20	0.18
C_c	0.07	0.08	0.18	0.10	0.15
B_c	0.14	0.16	0.05	0.19	0.09
D_c	0.04	0.06	0.17	0.07	0.08
PC_c	0.31	0.29	0.41	0.33	0.31
VR_c	0.32	0.33	0.39	0.35	0.34

Furthermore, one conjecture inferred from the experiment results (I-III) is that two-mode network properties (e.g., density, reciprocity, centralization) may affect the correlation among bipartite centrality measures, as well as their accuracy and performance. For instance, one observation from Table 4 and Figure 4 is that as network density increases, the correlation between Bi-face and closeness, eigenvector, and degree increases, while its correlation with betweenness decreases. This observation, however, does not clearly reflect the correlation between Bi-face and both percolation and vote-rank because *Women-Davis* has a lower density than *CoinToss* and *Dirichlet*, and Bi-face is more correlated with the two centrality measures on *Women-Davis* than on *CoinToss*. While this shows that network density influences how well different centrality measures correlate with one another, it also indicates that the network density is not the only factor and that other network properties may have an impact on such correlations. Since the study of the network properties is outside the scope of this paper, we could explore the effects of reciprocity and centralization on Bi-face in our future work.

V. CONCLUSION

The detection of influential nodes in a two-mode network is frequently an important task in scientific and industrial data analysis pipelines for explaining various behaviours and outcomes. Our work here addressed an obvious gap in the present CNA literature, namely the efficient identification of key nodes by combining both local cohesiveness and global network flow aspects of centrality through the use of FCA mathematical formalization. On this basis, we devised *Bi-face*, a new bipartite centrality measure that quantifies the prominence of a node in a two-mode network based on its presence in influential overlapping bicliques and bridges. While we focused on two-mode networks here, the approach can easily be modified to accommodate other complex network representations like multilayer networks.

From a conceptual perspective, the Bi-face score is a distinct centrality in the following three elements: (i) it uses the concept lattice formulation to efficiently extract overlapping bicliques and bridges, (ii) it leverages concept faces to refine bicliques from non-influential nodes and detect influential bridges, and (iii) it exploits the fact that influential bridges and overlapping bicliques with a large number of important neighbour nodes are likely to contain key central nodes. As a result, it measures how a node affects and is influenced by its important neighbours through refined bicliques, while also linking the network dense substructures via its existence in influential bridges. According to a thorough empirical study on several synthetic and real-life two-mode networks (see Section IV), the Bi-face score can identify key nodes more accurately and efficiently than other state-of-the-art centrality indices such as degree, betweenness, closeness, eigenvector, percolation, and vote-rank.

ACKNOWLEDGMENT

The authors thank the anonymous reviewers for their careful reading of the manuscript and their numerous insightful comments and suggestions.

REFERENCES

[1] 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.

[2] 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.

[3] S. P. Borgatti and M. G. Everett, "Network analysis of 2-mode data," *Social Netw.*, vol. 19, no. 3, pp. 243–269, 1997.

[4] 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, Aug. 2010.

[5] S. P. Borgatti and D. S. Halgin, "Analyzing affiliation networks," in *The Sage Handbook of Social Network Analysis*, vol. 1. Thousand Oaks, CA, USA: Sage, 2011, pp. 417–433.

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

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

[8] U. Brandes, S. P. Borgatti, and L. C. Freeman, "Maintaining the duality of closeness and betweenness centrality," *Social Netw.*, vol. 44, pp. 153–159, Jan. 2016.

[9] M. E. J. Newman, "A measure of betweenness centrality based on random walks," *Soc. Netw.*, vol. 27, no. 1, pp. 39–54, 2005.

[10] M. Newman, *Networks*. Oxford, U.K.: Oxford Univ. Press, 2018.

[11] M. Piraveenan, M. Prokopenko, and L. Hossain, "Percolation centrality: Quantifying graph-theoretic impact of nodes during percolation in networks," *PLoS ONE*, vol. 8, no. 1, Jan. 2013, Art. no. e53095.

[12] L. A. Meyers, "Contact network epidemiology: Bond percolation applied to infectious disease prediction and control," *Bull. Amer. Math. Soc.*, vol. 44, no. 1, pp. 63–86, 2007.

[13] 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, 2014, Art. no. 052808.

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

[15] 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.

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

[17] 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.

[18] Y. Bastide, N. Pasquier, R. Taouil, G. Stumme, and L. Lakhal, "Mining minimal non-redundant association rules using frequent closed itemsets," in *Proc. Int. Conf. Comput. Log.* Berlin, Germany: Springer, 2000, pp. 972–986.

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

[20] 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.

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

[22] M. H. Ibrahim, R. Missaoui, and J. Vaillancourt, "Cross-face centrality: A new measure for identifying key nodes in networks based on formal concept analysis," *IEEE Access*, vol. 8, pp. 206901–206913, 2020.

[23] C. Seierstad and T. Opsahl, "For the few not the many? The effects of affirmative action on presence, prominence, and social capital of women directors in Norway," *Scandin. J. Manage.*, vol. 27, no. 1, pp. 44–54, Mar. 2011.

[24] M. Morsey, J. Lehmann, S. Auer, C. Stadler, and S. Hellmann, "DBpedia and the live extraction of structured data from Wikipedia," *Program*, vol. 46, no. 2, pp. 157–181, Apr. 2012.

[25] S. P. Borgatti, "2-mode concepts in social network analysis," *Encyclopedia Complex. Syst. Sci.*, vol. 6, pp. 8279–8291, 2009.

[26] L. C. Freeman, "Finding social groups: A meta-analysis of the southern women data," in *Dynamic Social Network Modelling and Analysis: Workshop Summary and Papers*, K. C. R. Breiger and P. Pattison, Eds. Washington, DC, USA: National Academies Press, 2003, pp. 1–60.

[27] M. Felde, T. Hanika, and G. Stumme, "Null models for formal contexts," *Information*, vol. 11, no. 3, p. 135, Feb. 2020.

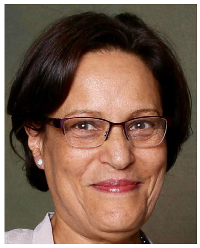
[28] M. Felde and T. Hanika, "Formal context generation using Dirichlet distributions," in *Proc. Int. Conf. Conceptual Struct.* Switzerland: Springer, 2019, pp. 57–71.

[29] J.-X. Zhang, D.-B. Chen, Q. Dong, and Z.-D. Zhao, "Identifying a set of influential spreaders in complex networks," *Sci. Rep.*, vol. 6, no. 1, p. 27823, 2016.

[30] 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.

[31] 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.

MOHAMED HAMZA IBRAHIM received the Ph.D. degree in computer engineering from the University of Montreal, Canada, in 2015. He is currently a Research Assistant 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 University of Montreal, in 2016, and the University of Quebec in Outaouais, in 2018. He also worked as a Data Scientist at 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 about 34 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 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 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, in 1987. He is currently a Affiliated Professor at HEC Montréal, after two full mandates as the President of the Université du Québec en Outaouais. His current research interests include theoretical studies of random media as well as knowledge discovery in large databases. He has served on numerous committees, notably as member for 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, as well as the Chair for the Mathematics and Statistics Group, Natural Sciences and Engineering Research Council of Canada, from 2003 to 2006.