IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 18, 2021, accepted May 23, 2021, date of publication June 3, 2021, date of current version June 14, 2021.

*Digital Object Identifier 10.1109/ACCESS.2021.3086027*

# A Group-Based Centrality for Undirected Multiplex Networks: A Case Study of the Brazilian Car Wash Operation

**BRUNO CÉSAR BARRETO DE FIGUEIRÊDO, FABÍOLA GUERRA NAKAMURA, AND EDUARDO FREIRE NAKAMURA**

ICOMP, Federal University of Amazonas, Campus Universitário Senador Arthur Virgílio Filho, Manaus 69067-005, Brazil

Corresponding author: Bruno César Barreto de Figueirêdo (brunocesar@tce.rr.leg.br)

This work was supported in part by the Institute of Computing, Federal University of Amazonas (UFAM), Manaus, AM–Brazil, in part by the Foundation for Research Support of the State of Amazonas (FAPEAM)—POSGRAD 2020 (Resolution 002/2020), and in part by the Coordination for the Improvement of Higher Education Personnel—Brazil (CAPES)—Finance Code 001.

**ABSTRACT** One challenging issue in information science, biological systems, and many other fields is determining the most central or relevant networked systems agents. These networks usually describe scenarios using nodes (objects) and edges (the objects' relations). The so-called standard centrality measures aim to solve this kind of challenge, ranking the nodes by their supposed relevance and elect the most relevant nodes. This problem becomes more challenging when one single network is not enough to depict the whole scenario. In these cases, we can work with multiplex networks characterized by a set of network layers, each describing interrelationships that can change depending on external factors, e.g., time. This paper proposes a new centrality measure, the Group-based Centrality for Undirected Multiplex Networks, to find the most relevant nodes in an undirected multiplex network. As a case study, we use a Brazilian corruption investigation known as the Car Wash Operation. Our proposed centrality outperforms well-known centrality methods such as betweenness, eigenvector, weighted degree, Multiplex PageRank, closeness, and cross-layer degree centrality.

**INDEX TERMS** Group-based centrality measures, multiplex centrality measures, multiplex networks.

## I. INTRODUCTION

Imagine a large number of individuals (nodes) interacting in a network. Somewhat ranking these individuals is a challenge, given that those interactions can vary in a non-predictable way. The centrality measures have this challenging mission. We can understand the term centrality as a tool to quantify the relevance of nodes in a network [1]. The study of centrality measures began in the '50s, introducing the role of nodes in communications patterns [2]. Since then, constant studies aim to improve ranking results. We obtain these methods through specific case considerations about the way social interactions function, mainly based on inferences about the spread of information across a group [3]. Examples of standard centrality metrics include the betweenness [4], [5], eigenvector centrality [6], PageRank [7], and weighted degree [8].

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

Typically, these metrics use the nodes' topological position as the basis for its ranking, e.g., the number of node connections, the connections of node neighbors, the number of walks, and paths going across the node. Different metrics try to provide an answer to the question: "*which are the most important nodes in a network?*" [9], [10]. The range of applications to this technology is vast, e.g., epidemiology [9], [11], [12], economics [13], [14], neuro-sciences [15], engineering [16], and fraud detection [17]. However, tests on a broad set of networks show intrinsic highlighting limitations of standard centrality metrics for finding the most relevant nodes in complex networks [18].

One single complex network cannot ideally describe some natural systems. Let us consider an extensive collection of books containing a single main story; some characters emerge as very important during the plot, but they lose relevance throughout the story. How to aim for a fair ranking once the story's nuances change the importance of characters as

it continues throughout multiple books? What would be the solution for modeling this real situation in a single complex network? How can conventional centrality metrics provide a ranking of nodes in situations like this?

In these cases, a natural solution is to use networks that model the real situation [19]. Considering the example and supposing that each book is modeled as a separate complex network, obtaining the ranking of the whole book collection, it would be necessary to add the networks that model each book as a single network and only then use a standard centrality metric to obtain the most relevant nodes.

However, this approach is flawed since various networks can have singular characteristics that change with time. Now let us consider a political dispute scenario between democrats and republicans, in which the nodes are the deputies and the edges the meetings between them. As a metric, let the most relevant deputies have more meetings with the opposite political party's deputies. Four networks describe this scenario, each one being a year of appointments. Now imagine that, for the first two years, a restricted group of five congresspeople of both parties had an enormous number of meetings, and, after a political rupture, these meetings stopped. In this scenario, if we unite the four networks in a single network, as these five congresspersons had an enormous number of meetings in the first two years, they can appear, mistakenly, as the five most relevant nodes of the whole scenario. It would be a misinterpretation once they did not have this status throughout the entire period. In other words, the combination of all networks that describe a scenario in one network can lead us to an incorrect interpretation of the facts that occur in different scenarios over time.

Since a single network cannot represent all scenarios, the multilayer networks attract growing interested and research once they can describe multiple types of interactions between any pair of nodes [20]. Considering the example above, we can divide the four years of meetings into four layers, and have a fairer nodes ranking, thinking of each year as an independent part of a scenario. However, standard centralities don't work with this network structure. Thus, proposals of centrality metrics, including the metrics that aim to work with multilayered networks, emerge as the natural extension of standard metrics [21]–[23]; being, in some cases, an improvement of the original algorithms [22], [23]. In our paper, we work with the multiplex networks, a particular type of multilayer networks, described in Section (IV).

We propose the novel GCMN centrality measure for situations requiring a multilayered complex network. GCMN allows a ranking that considers the node's existence in the different layers of a multiplex network in a hierarchical way. It has a ranking based on group formatting according to the node's weight into the layers — dropping the mistaken interpretation mentioned above. GCMN also offers an alternative approach for the centrality metrics defined by the random walk in undirected multilayered networks [22], [23], proposing a novel strategy based on the intersection of the layers as a parameter for assigning weight to nodes. For GCMN centrality formalization (Section IV) we use the tensorial notation (Section III-A) and the multiplex networks theory (Section III-B).

The main contributions of our work are: (i) the proposal of a centrality measure with a novel ranking approach, based on node's hierarchical grouping according to their weights (Section IV); (ii) an equivalent to or better performance than other known centralities, measured with Accuracy, Precision, Recall, and $F_1$ Score (Section VI-A); and a metric proposed in this paper (Section V-A).

We organize the remainder of the paper as follows. In Section II we discuss the related work. In Section III-A we briefly describe the tensorial notation, defined in [24], adopted throughout the paper. In Section III-B we use this notation to describe the Undirected Multiplex Networks. In Section IV we propose and formalize the GCMN Centrality. In Section V, we describe the methodology, in particular, the application of the GCMN centrality in our case study and, in Subsection V-A, we propose a metric to evaluate its performance. In Section VI we discuss the results and benefits of the GCMN Centrality. Finally, we discuss our findings and future work in Section VII.

## II. RELATED WORK

Our article addresses a particular type of network called multiplex networks: a particular type of multilayer network in which each node appears in different layers, and each layer describes all the edges of a given type. These nodes cannot have connections with other nodes in other layers. In this paper, we represent multiplex networks as a three-dimensional matrix of size $(V \times V) \times L$, in which $V$ represents the vertices (nodes), and $L$ the layers, or dimensions [25], [26]. The use of standard centrality measures [4]–[8] had to be revised to cover the multiple layers of this new network structure.

This review led to the proposition of extensions of the standard centrality measures as being a natural path to be followed. Some proposals for adapting these centrality measures have emerged, such as "Novel Multiplex PageRank in Multilayer Networks" [22], "Random walk centrality in interconnected multilayer networks" [23] and "Random Walks on Multiplex Networks: Supplementary Information for Navigability of Interconnected Networks under Random Failures" [27]. Therefore, this type of extension of the standard centralities, although valid, does not bring new ranking strategies.

Recent strategies have emerged with a specific focus on these new multilayer network structures, like The CLDC [21]. In the CLDC centrality measure, the ranking of node $x$ is calculated as a ratio between the number of nodes connected with node $x$ and the total number of all nodes in the network (decreased by one). Thus, the CLDC cross-layer centrality computed as a sum of both incoming and outgoing edge weights from node $x$ towards its multilayered neighborhood divided by the number of layers and the total number of network members.

We also find strategy proposals for specific goals like classifying nodes in urban mobility networks [28]. For such a use-case, the authors locate places in a city by designing weighted directed graphs whose nodes denote city locations, and weighted edges represent the number of trips between them. The nodes' attributes indicate socio-economic characteristics at a particular location in the city, and combines this information with "hotposts" of different types of socio-economic activities.

Centrality measures are not suited for multiple criteria decision scenarios. Group-based strategies are a viable option in these situations. A case study presents a group-based approach based on the weighted $k$-means to rank venture capital firms in the Chinese investment market as an alternative strategy [29]. Previous authors propose a way to generalize block-modeling for hierarchical decomposition, using the $k$-means method to decompose a social network into groups of nodes having the existence of congruent profiles of dissimilarities with other nodes as a criterion [30]. Therefore, group-based proposals with an associated hierarchy classify nodes according to their relevance, sorting out the nodes present in the core of the networks as being of most interest in the research from the network's periphery associated nodes, with less interest [31].

GCMN centrality combines aspects of various centrality measures. It is a new strategy for ranking nodes in multilayer networks based on hierarchical node grouping. Thus, the GCMN centrality elects the most relevant nodes of a multiplex network in multiple criteria decision scenarios.

## III. PRELIMINARIES

### A. TENSORIAL NOTATION

In our paper, we will use the tensorial notation representing adjacent matrices using higher-order algebra [24]. A significant advantage of using tensors' formalism relies on its compactness. We can write an adjacency matrix, or tensor, by using a compact notation that is very useful for the generalization of network descriptors to multilayer or, in our specific case, multiplex networks (Section III-B).

In tensorial notation, a row vector $i \in \mathbb{N}$ is given by a co-variant vector $i_\alpha(\alpha = [1, N])$. Its corresponding contravariant vector $i^\alpha$ (i.e., its dual vector) is a column vector in Euclidean space. A canonical vector is assigned to each node and a mixed rank-4 adjacency tensor represents the corresponding interconnected multilayer network. In this case, a tensor $A_{ij}^{\alpha\beta}$ can represent the intensity of the relationship (which may not be symmetric) from a node $i$ in layer $\alpha$ to a node $j$ in layer $\beta$.

To formalize the GCMN centrality, we will use an intra-layer adjacency tensor for the 2nd-order tensor $A_{ij}^{\alpha}$. This kind of tensor indicates the relationships between nodes $i$ and $j$ within the same layer $\alpha$. In section III-B, we use tensorial notation for a fictional network.

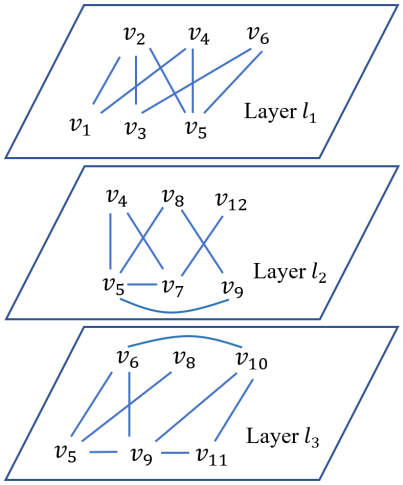



**FIGURE 1.** An example of a multiplex undirected network.

### B. UNDIRECTED MULTIPLEX NETWORKS

Multiplex networks are a particular type of multilayer networks in which each node appears in different layers, and each layer describes all the edges of a given type. These nodes cannot have connections with other nodes in other layers. A three-dimensional matrix of size $(V \times V) \times L$, in which $V$ represents the vertices (nodes), and $L$ the layers, or dimensions; with entries, $A_{ij}^{\alpha}$, is enough to represent the structure of the system [25], [26]. By using tensorial notation [24], adjacency matrices are indicated by multiplex adjacency tensors $A_{ij}^{\alpha}$ to encode connections between nodes $\{i, j \mid i, j \in V\}$ in layer $\{\alpha \mid \alpha \in L\}$.

A tensor is an algebraic object that describes a multi-linear relationship between sets of algebraic objects. In our case, these objects are the adjacency matrices indicated by the nodes in $V$ and the layers in $L$. The $A_{ij}^{\alpha}$ values will be one when we have an edge between the nodes $i$ and $j$ in the layer $\alpha$, and zero otherwise.

Thus, we can represent an UMN (Undirected Multiplex Network) as a triplet $(V, E, L)$, with $V$ being the nodes, $L$ the layers, and $E$ a set of entries $\{A_{ij}^{\alpha} \mid i, j \in V,\ \alpha \in L, i \neq j\}$; in which, for any two entries $\{A_{ij}^{\alpha}, A_{xy}^{\beta} \in E\}$, if $i = x$ and $j = y$, then $\alpha \neq \beta$. Note than we are dealing with undirected networks, so an entry $A_{ij}^{\alpha}$ is equivalent to $A_{ji}^{\alpha}$ [21].

Fig. 1 brings an example of a UMN in which: $V = \{v_1, \ldots v_{12}\}$, $L = \{l_1, l_2, l_3\}$ and $E$ is a set of adjacency matrices representing the connections between the nodes in $V$ in the layers in $L$. Every adjacency matrix $A_{ij}^{l_1}$ represents the connections between the nodes $\{i, j \mid i, j \in V\}$ in the layer $l_1$, with the components equal to one when there is an edge between $i$ and $j$, and zero otherwise.

## IV. THE GROUP-BASED CENTRALITY FOR MULTIPLEX NETWORKS – GCMN

GCMN's initial premise is the existence of a network with multiple layers, and, inside these layers, we have nodes interconnected by undirected edges. Nodes represent elements (e.g., people, companies), and undirected edges represent

their relationships (e.g., friendship, contracts). The layers represent the different contexts in which these elements may or may not be related to each other. The set of network layers represents the whole scenario of the analysis. Note that, as we deal with multiplex networks, edges can have different types in each layer. Thus, the GCMN strategy is firmly based on the network topology, considering the presence of nodes in each layer and, as a second-ranking criterion, the degree of each node. The centrality does not explore any other information inserted in a node, nor the direction of connections between nodes.

An application of networks with multiple layers is the friendship relationships in social networks (e.g., Facebook, Instagram, or Twitter). In this scenario, each layer maps the relationships in a social network in which individuals can relate to each other. This way, each layer can have the same individuals (nodes) with different connections (edges).

As the GCMN centrality works with undirected multiplex networks (*UMN*), we specify the triple $(V, E, L)$, with the set $V$, of nodes, and the set $L$, of layers. Thus, we define the set $E$ (of edges) as the relations between the nodes in $V$ in each layer of $L$, as:

$$E = \{A_{ij}^{\alpha} \mid i, j \in V \wedge \alpha \in L \wedge i \neq j\}. \tag{1}$$

The GCMN centrality splits the nodes into groups. The criterion for this grouping is the number of layers $\alpha \in L$ that a node $\{i \mid A_{ij}^{\alpha} \in E\}$, is present. We call this property the weight $W$ of a node. Thus, the definition of the $W$ function is:

$$W_i = \mid \{\alpha \mid A_{ij}^{\alpha} \in E\} \mid. \tag{2}$$

We define the group $G$ of nodes $i$, which have the same weight, as:

$$G_w = \{i \mid W_i = w\}, \tag{3}$$

in which $w$ is the weight of the nodes into group $G_w$. Our premise is that the weight $w$ of a node is proportional to its importance, and, as a consequence, a group $G_w$ should have more relevant nodes than a group $G_{w-1}$. We only consider groups with weight $w \geq 2$ e.g. groups containing nodes that appear in at least two layers.

Another essential concept is the degree $D$ of a node, which is the sum of its connected edges throughout all layers [32]. Thus, the degree of a node is:

$$D_i = \sum_{\alpha=1}^{L} A_{ij}^{\alpha} \tag{4}$$

in which the tensor $A_{ij}^{\alpha}$ has all components equal to one. The GCMN centrality ranking considers two criteria simultaneously: the group $G$ of a node $i$ and its degree $D$. Our ranking considers the group as the first criterion and the degree as the second. This way, nodes that appear only in a small subset of layers will not rank as significant even with a large degree. Thus, solving the problem of distortion found in the evaluation of the most relevant nodes in a scenario composed of multiple complex networks, seen in section I.

For a node $i$ to be considered as relevant, it must satisfy two simultaneous criteria: be present in a significant number of layers, maximizing $W_i$, and also have a significant degree $D_i$. Since the first criterion overlaps with the second, a node that appears in a small subset of layers cannot be considered as relevant in the context as a whole.

It is necessary to ensure that no node in a group $G_w$ ranks higher than any other of a group $G_{w-1}$. We achieve this by defining the general formulation of the ranking $R$ of a node $i$ as:

$$R_i = \varphi(W_i) + D_i, \tag{5}$$

in which $\varphi$ must ensure that $W_i > W_j \rightarrow R_i > R_j$. Note that the $\varphi$ function returns the same value for all the nodes in a group, considering that these nodes have the same weight $W$ (Equation 3), so the degree $D$ ranks these nodes into their groups.

The minimum of the $\varphi$ function occurs when $W_i = 2$, since the GCMN centrality ranking considers nodes with associated weight, at least, equal to two. Considering Equation 5, to guarantee that $W_i > W_j \rightarrow R_i > R_j$, we have that $\varphi(2) > \max_{z \in V} D_z$. That is, the $\varphi$ function must guarantee that, in its worst case, $\varphi$ exceeds the highest degree $D$.

Thus, the model allows us to define any $\varphi$ function, as long as the above condition is respected. Considering that $D_z \in \mathbb{N}$, and that $W_i \geq 2$, we have that $\max_{z \in V} D_z + 1$ is the lowest possible value for $\varphi$. Therefore, we propose the $\varphi$ function as

$$\varphi(W_i) = \left(\max_{z \in V} D_z + 1\right)(W_i - 1). \tag{6}$$

So, as the minimum of $\varphi$ occurs when $W_i = 2$, we have that $\varphi(2) = (\max_{z \in V} D_z + 1)(2 - 1) > \max_{z \in V} D_z$. Notice that this is our proposal for the $\varphi$ function, which we will use in our case study; other $\varphi$ functions are also valid as long as they respect the condition $W_i > W_j \rightarrow R_i > R_j$.

As an example, we will now apply the GCMN centrality to the multiplex network proposed in Fig. 1. In Table 1, we have the application of functions: $W$, $\varphi$, $D$, and $R$ (Equations 2, 4, 5, and 6) for all nodes in $V$ and all layers in $L$. As the $\max_{i \in V} D_i = 10$, for the node $v_5$, we have that $\varphi(W_i) = (10 + 1) \cdot (W_i - 1)$ (Equation 6), and two groups: $G_2 = \{v_4, v_6, v_8, v_9\}$ and $G_3 = \{v_5\}$ (Equation 3).

The final ranking $R$ (or its normalized version $\hat{R}$) shows the node $v_5$ as the first placed, followed by $v_9$, $v_6$, $v_4$, and $v_8$. This is the expected result, since $W_{v5}$ reached the greatest value for the nodes in $V$. As the other nodes have the same weight $W$, their ranks are based on their degrees $D$. As the nodes $v_1$, $v_2$, $v_3$, $v_7$, $v_{10}$, $v_{11}$, and $v_{12}$ have weight equal to one; according to the GCMN centrality, they are not relevant, and will not have a ranking.

Since GCMN works with groups composed of the nodes present in each layer, an extreme case would be that all nodes were present in all layers. In this case, the hierarchy of groups would not be applicable, and all nodes would be in the same group. Therefore, the only criterion for classification would

**TABLE 1.** GCMN centrality application example.

| | $v_1$ | $v_2$ | $v_3$ | $v_4$ | $v_5$ | $v_6$ | $v_7$ | $v_8$ | $v_9$ | $v_{10}$ | $v_{11}$ | $v_{12}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $W$ | 1 | 1 | 1 | 2 | 3 | 2 | 1 | 2 | 2 | 1 | 1 | 1 |
| $\varphi$ | | | | 11 | 22 | 22 | | 11 | 11 | | | |
| $D$ | | | | 4 | 10 | 5 | | 3 | 6 | | | |
| $R$ | | | | 15 | 32 | 16 | | 14 | 17 | | | |
| $\hat{R}$ | | | | 0.468 | 1.000 | 0.500 | | 0.437 | 0.531 | | | |

**Algorithm 1** Construction of the GCMN Centrality Ranking

**Result**: $X$ // set with the ranked nodes
Input: $N : (V, E, L)$;
Let $X = \{\}$;
**for** $i$ *in* $N.V$ **do**
  Let $x$:(node,weight, degree);
  $x$.node $= i$ ;
  $x$.weight $=$ W($i$) //*Equation* 2;
  $x$.degree $=$ D($i$) //*Equation* 4;
  $X = X \cup x$ ;
**end**

**Algorithm 2** Retrieving a Node's Ranking

**Result**: $r \in \mathbb{N}$ // the rank for the node $i$
Input: $i$, $X$:{(node, weight, degree)};
Let $x$ in $X \mid i = x$.node;
$r = \varphi(x.\text{weight}) + x.\text{degree}$ // Equations 5 and 6;

be the degree of the node, which would lead us to a poor ranking.

Another case to be clarified is that of isolated nodes without connections to other nodes in a layer. These nodes are considered for classification by the GCMN. However, we assume that the probability that a node participates in isolation in a significant group of layers is minimal.

Finally, there is the case of nodes with self-connections. In this case, these connections are equivalent to two regular links.

### A. ALGORITHM AND COMPLEXITY

The algorithm's input is a multiplex complex network $N : (V, E, L)$, and the output is a set $X$ with all nodes' ranking. To make the algorithm easier to read, we consider that the network $N$ is visible for the $W$ and $D$ functions. Algorithm 1 constructs the GCMN centrality ranking. This procedure performs primitive operations with complexity equal to $O(1)$; an external loop over all vertices in $V$, with two independent internal loops over all edges in $E$, when calling Equations 2 and 4. Therefore, Algorithm 1 runs in $O(|V||E|)$.

The algorithm input is the node $i$ for which we want the ranking and an $X$ set of nodes with their rankings. This $X$ set is the output of Algorithm 1. Algorithm 2 retrieves a node given its index. Assuming a hashing data structure for $X$ and our proposed $\varphi$, we can compute all $D_z \mid z \in V$ needed for $\varphi$ while populating $X$. Therefore, we needn't traverse the network again in Algorithm 2, essentially allowing ranking retrieval in $O(1)$.

## V. MATERIALS AND METHODS

We illustrate our proposal for the GCMN centrality with a case study analysis. The chosen case was a Brazilian corruption investigation, called the Car Wash Operation, started in 2009 by the Brazilian Federal Police, which investigates the practice of financial crimes and embezzlement of public funds. To encourage criminals to collaborate with investigations, the Brazilian Federal Prosecutor Office signed leniency agreements. In exchange for criminal information, the convicted criminals may have their sentences reduced or even extinct. In this context, statement information is vital in identifying influential entities (individuals), helping the law enforcement authorities direct their investigations. Our case study is about five testimonies offered by criminals convicted by the Car Wash Operation [19]. These testimonies explain in detail the corruption mechanism that exists in Brazil's highest's political levels. The convicted individuals had their sentences reduced or had other benefits like serving the prison sentence at home as a reward for collaborating with justice.

Since the GCMN centrality deals with multiplex networks, the natural choice was to split the five testimonies into five layers ($L$), in which the nodes ($V$) refers to the individuals that cited by each convicted, and the edges ($E$), the joint occurrence of these individuals in the testimonies' excerpts (Equation 1).

The next step was to determine the weight ($W$) of each node (Equation 2), and create the groups ($G$) of nodes (Equation 3) according to their weights. As a result, we achieve $\mid G_2 \mid = 62$, $\mid G_3 \mid = 29$, $\mid G_4 \mid = 16$, and $\mid G_5 \mid = 5$. Note that, as we have five layers, the number of groups considered as relevant is four, that is, groups with nodes $i$ whose weight is $w_i \geq 2$. The last step was apply the Equations 4, 5, and 6, to all nodes in groups $G_2$ to $G_5$ to obtain the GCMN centrality rank. In our case study, the $\max_{z \in V} D_z = 95$, so, our $\varphi$ function will be $\varphi(W_i) = (95 + 1) \,.\, (W_i - 1)$ (Equation 6).

The following subsections compare the GCMN centrality ranking with the ranks of the standard centrality measures: weighted degree (WD) [8], betweenness (BW) [4], [5], eigenvector (EV) [6], and closeness (CL) [4]; the ANN SCORE which represents the normalized geometric mean of the metrics (BW, EV, and WD) [19]; and two multilayered centrality measure: Cross-layer degree centrality (CLDC) [21] and The Multiplex PageRank (MPR) [22].

### A. PROPOSED EVALUATION METRICS

GCMN is a group-based centrality that divides nodes into groups according to their relevance. The other centrality measures we compare GCMN with do not have this concept. This way, for comparison purposes, we will consider the number of nodes in each group and the same number of nodes classified by the other centrality measures, preserving their ranking. Table 2 brings this process description, being $i$ the highest reached the weight and $j$ the number of compared centrality measures.

Applying the ranking to the case study, we have that, for group $G_5$, with five nodes, we consider the first five best-classified nodes for all other centrality measures. In this way, it is possible to compare group $G_5$ with the best-ranked nodes in all other centrality measures. The group $G_4$ has sixteen nodes, and we consider the nodes classified between six and twenty-one in all other measures. This process is the same for groups $G_3$ and $G_2$.

For analysis purposes, a relevant parameter is the gain of the company. Therefore, the proposal for a metric that considers values a parameter of relevance for companies is coherent. This way, we will measure the groups' relevance according to the values obtained by the companies belonging to their groups, checking if the most relevant groups were able to point out the companies with the most significant gains. Remembering that our premise is that the scalar value that indicates the weight ($W_i$) of a node $i$ (Equation 2) is proportional to its importance, and, as a consequence, a group $G_w$ should have more relevant nodes than a group $G_{w-1}$ (Equation 3).

In the Car Wash scenario, we will consider five possible values for a legal status $S$ of a node $v$, as $S(v)$ where

$$S : s \rightarrow \exists! s | s \in \{NotInvestigated, Investigated, Denounced, Defendant, Convicted\}, \quad (7)$$

each one with its characteristic and degree of importance.

*NotInvestigated*: there was no investigation, or the investigation concluded that the individual was innocent, being the minor classification of our scale and encompasses the so-called false positives;

*Investigated*: there is an ongoing investigation, but with no result yet. The analysis cannot consider these individuals;

*Denounced*: the prosecutor lodges a formal complaint and the individuals are relevant for our analysis;

*Defendant*: there was the acceptance of the complaint, and the individual will go to trial for some crime related to Car Wash Operation. This individual has high relevance for our analysis;

*Convicted*: There was a formal trial, and the individual received a sentence for crimes related to Car Wash Operation, being the most relevant group to our analysis.

According to the legal rite of an individual's indictment process, from his investigation to his eventual conviction, we find that situations *Not Investigated / Acquitted*, *Investigated* refer only to police suspicions; *Denounced*, *Defendant*, *Convicted* involve participation of a prosecutor and/or a judge. Thus, we understand as reasonable to divide the five situations into two distinct sets: most relevant legal status — $MRS = \{Denounced, Defendant, Convicted\}$ — and least relevant legal status — $LRS = \{NotInvestigated/Acquitted, Investigated\}$. So, the relevance of the node is *MRS* or *LRS*, according to its legal status, formally defined as $S(v)$.

We defined a metric called Relevance Index of a Group to determine the accuracy of the results of a group as

$$RI(G(n)) = \left( \frac{|\{v_i | S(v_i) \in MRS \wedge v_i \in G(n)\}|}{|G(n)|} \times 100 \right), \quad (8)$$

which corresponds to the percentage of nodes of that group whose legal status belongs to MRS, where $G(n)$ is the group under analysis, and $v_i$ is a node that belongs to $G(n)$.

As an extension of the relevance index we compute the general relevance

$$GR = \frac{|\{v_i | S(v_i) \in MRS|}{|\{v_j | S(v_j) \in (MRS \cup LRS)\}|}, \quad (9)$$

that is, the sum of nodes on MRS from $G_2$ to $G_5$, divided by the total of nodes, that is $MRS \cup LRS$.

This approach intends to show the GCMN's centrality performance in determining the most relevant nodes in a set of networks, and the weight's effectiveness as a nodes classification criterion, demonstrating that the weight's growth is directly proportional to the relevance of the nodes selected. We will now compare the results of applying the GCMN centrality with well-established centralities [19] showing that the GCMN centrality can point out the most relevant nodes in a more effective way than these centrality measures.

## VI. RESULTS AND DISCUSSION

The results and its discussion will consider two parameters: the use of known metrics such as Accuracy, Precision, Recall and $F_1$ Score (Subsection VI-A), and the use of our proposed metric (Subsection V-A) to evaluate the GCMN performance in three aspects — "The weight as a grouping parameter" (Subsection VI-B), "The Relevance per group" (Subsections VI-C and VI-D), and "Qualitative analysis of groups $G_3$ to $G_5$" (Subsection VI-E).

### A. ACCURACY, PRECISION, RECALL AND $F_1$ SCORE

One typical way to quantify the quality of classification and clustering tasks is using Precision and Recall metrics. Precision corresponds to the fraction of relevant elements among the retrieved elements, while Recall evaluates the fraction of the total amount of relevant elements that were retrieved. Both metrics help to measure the relevance of the ranked nodes. The $F_1$ score is the harmonic mean of Precision and Recall and summarizes the quality of clustering in a value.

To calculate the Precision and the Recall, we have to divide our universe of elements into four groups; we have $TP-truepositives$ (detected correctly), $FP-falsepositives$ (detected incorrectly), $FN-falsenegatives$ (not detected

**TABLE 2.** Nodes distribution from groups to centrality measures.

| Group $G(i)$ | | | | |
|---|---|---|---|---|
| | Centrality 1 | Centrality 2 | ... | Centrality $j$ |
| Node 1 | ranking for node 1 | ranking for node 1 | ... | ranking for node 1 |
| ... | ... | ... | ... | ... |
| Node $n$ | ranking for node $n$ | ranking for node $n$ | ... | ranking for node $n$ |
| Group $G(i-1)$ | | | | |
| | Centrality 1 | Centrality 2 | ... | Centrality $j$ |
| Node $n+1$ | ranking for node $n+1$ | ranking for node $n+1$ | ... | ranking for node $n+1$ |
| ... | ... | ... | ... | ... |

**TABLE 3.** Individuals grouping for Precision and Recall analysis.

| | $MRS$ | $LRS$ |
|---|---|---|
| Detected ($G_3$ to $G_5$) | $truepositives$ | $falsepositives$ |
| Not Detected ($G_2$) | $falsenegatives$ | $truenegatives$ |

incorrectly), and *TP−truenegatives* (not detected correctly). The Precision is given by

$$precision = TP/(TP + FP) \tag{10}$$

and the Recall by

$$recall = TP/(TP + FN). \tag{11}$$

Taking our case study, we consider the detected individuals in *MRS* as *thetruepositives*; the detected individuals in *LRS* as *thefalsepositives*; the not detected in *MRS* as *thefalsenegatives*; and the not detected in *LRS* as *thetruenegatives* (Table 3).

The Precision analysis shows the percentage of relevant individuals, considering just the detected ones, was reached. We assume the individuals in *MRS* as the relevant nodes. It can lead us to which centrality we can use in a real situation, like an auditing process, e.g. In other words, which centrality can lead us to the most significant number of individuals to audit and find irregularities and the smallest number of individuals in which the audit process will not lead to any results. The GCMN and the Weighted Degree (WD) reached 90% Precision, being the two best-posicioned centralities.

The Recall analysis shows which centrality can point to a more significant number of relevant individuals, considering all individuals that should be detected. As the Precision analysis, it is useful for an investigation team when choosing which centrality to use. Thus, the GCMN centrality reached 51% Recall, which is a good result (Fig. 2).

The $F_1$ analysis brings an overview of the two metrics performance (Precision and Recall), showing that the GCMN centrality reached an overall good result, achieving 65%. (Fig. 2).

In our case study, accuracy refers to the degree of conformity of a calculated quantity to an actual value. Accuracy is closely related to precision, but it's not a synonym. A result is said to be accurate when it matches a particular target. In our case study, the target is the individuals with legal status with a high degree of severity. In this matter, we reached a high degree of precision (90%) with high accuracy (79%), meaning that the GCMN centrality ranking could point to the individuals in a precise way, reaching the goal with high accuracy, being the best performance among all the other centrality measures (Fig. 2). The Accuracy is given by

$$Accuracy = (TP + TN)/(TP + TN + FP + FN). \tag{12}$$

### B. THE WEIGHT AS A GROUPING PARAMETER

According to the GCMN centrality grouping, Fig. 3 shows individuals' normalized distribution per legal status and group. We realize that the number of individuals in the most relevant legal status is consistent with the groups' degree of relevance $G_2$ to $G_5$.

Analyzing the results of each group, considering the relevance index (Equation 8), we have:

$G_2$: this group shows 18 individuals in *LRS* \$*NotInvestigated*/ *Acquitted* and *Investigated*) and 44 in *MRS* (*Denounced*, *Defendant*, and *Convicted*). The relevance index of the group is 71%;

$G_3$: applying the same criteria as in the previous group, we found a relevance index of 86%. It is important to emphasize that there are only four individuals with the least relevant status (*LRS*);

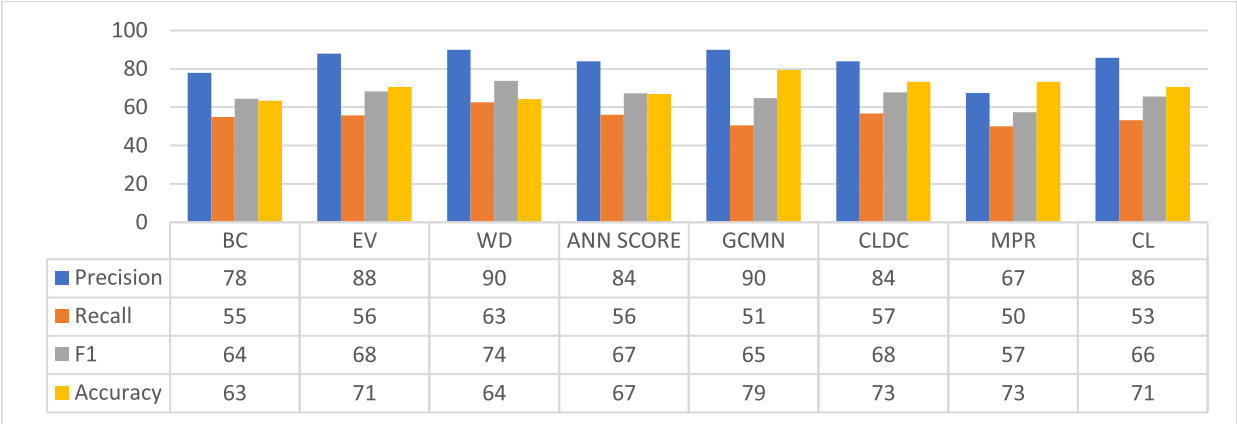
$G_4$: the relevance index grows once again to 94%. In this group, the inexistence of individuals with status investigated or denounced should be stressed. So, except by one individual, all the elements of the group have the status *Defendant* or *Convicted*;

$G_5$: the group's relevance index reaches the maximum percentage of 100%. It means that all the elements found are relevant.

After this analysis, we verify that the weight growth associated with the nodes is directly proportional to the severity related to their legal status. This fact shows that weight is an excellent choice as a parameter to build groups.

### C. RELEVANCE PER GROUP

Fig 4 compares the distribution of nodes per group and legal status, considering the GCMN centrality and the other




| | BC | EV | WD | ANN SCORE | GCMN | CLDC | MPR | CL |
|---|---|---|---|---|---|---|---|---|
| Precision | 78 | 88 | 90 | 84 | 90 | 84 | 67 | 86 |
| Recall | 55 | 56 | 63 | 56 | 51 | 57 | 50 | 53 |
| F1 | 64 | 68 | 74 | 67 | 65 | 68 | 57 | 66 |
| Accuracy | 63 | 71 | 64 | 67 | 79 | 73 | 73 | 71 |

**FIGURE 2.** Accuracy Precision Recall and $F_1$ comparative analysis.

centralities present in Almeida *et al.* (2017), in addition to three other centralities (CLDC, PR, and CL).

Taking the number of nodes of each group ($G_2$ to $G_5$) and analyzing them by the relevance criterion (*LRS* and *MRS*), we realize that the GCMN centrality has a general superior performance when compared with the other centrality measures. The GCMN centrality reached the best results on groups $G_2$ and $G_5$, ties with the centralities BC, WD, and ANN SCORE in the $G_4$ group, and lies behind the measures EV and WD, by just one individual, in the group $G_3$ (Table 4).

The GCMN centrality reached the best performance on the relevance index (Equation 8) on the groups $G_2$, $G_4$, and $G_5$; and the second best on group $G_3$. The GCMN centrality also reached superior general relevance i.e. 80 against 68 of the second best ranked metric, the CL (Equation 9) (Table 5). As a grouping criterion, the use of weight was influential not only for the GCMN centrality but also for other centralities whose relevance was, in general, consistent with this approach (Table 5). The only significant exception to this rule was the $G_5$ group. However, we should consider the reduced number of nodes in this group, which results in a higher significance on the relevance index by one element.

It is also important to note that, in the group $G_5$, the results of the GCMN centrality are consistent since there are no nodes related to the two less relevant status (*NotInvestigated*/*Acquitted* and *Investigated*), what occurs in BC, EV, WD, ANN SCORE, and CLDC.

### D. RELEVANCE PER GROUP, A CUMULATIVE ANALYSIS

Fig. 5 brings the distribution of nodes per group and status cumulatively. This way, the group $G_5$ is the same in Fig. 4, and, for the other groups, we can see the increase of relevance in a gradual, cumulative view, providing a more productive analysis and an overview of the results. As in Section VI-C, there will be a comparison between the centrality measures and the GCMN centrality, with the results highlighted the in Table 6.

For all groups, the number of individuals of the GCMN centrality in the *LRS* is the smallest, and, consequently, the number of nodes in the *MRS* is the highest (Table 6).

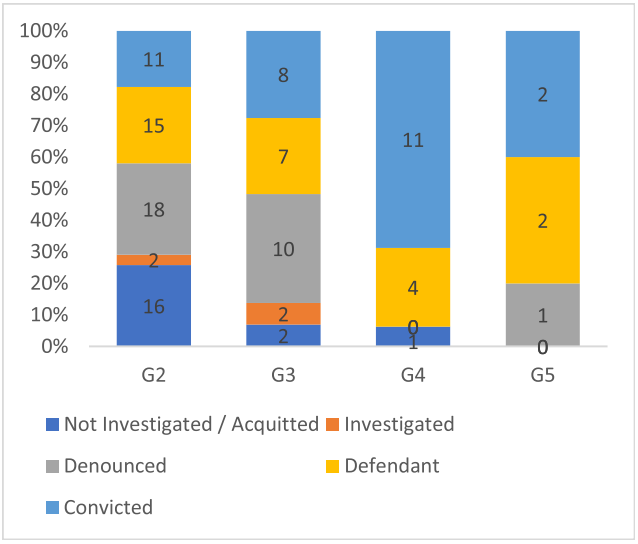



**FIGURE 3.** Illustration of the weight as a parameter to define groups.

The GCMN centrality presents the highest number of nodes associated with the most relevant legal status (*convicted*) from group $G_2$ to $G_4$. Despite group $G_5$ be an exception, it does not consist in a problem or deficiency considering that the group only has five nodes and that, in this kind of status, only one element is equivalent to 20% of the whole group, thus reducing the possibility of a more precise statistical analysis.

### E. QUALITATIVE ANALYSIS OF GROUPS $G_3$ TO $G_5$

The qualitative analysis intends to verify the ability of each metric to point out "novelties". This analysis is particularly useful for the discovery of the hitherto "unreachable" nodes.

Table 7 shows that the GCMN centrality could point out seventeen nodes not detected by the other centrality measures, that is, 42% more than the second-placed, the CLDC Centrality, being thirteen of the seventeen nodes pointed out in the most relevant legal status (*MRS*). Comparing this result with all the other centralities (fifty-five nodes), we verify that the GCMN centrality pointed to 31% of the "unreachable" and relevant nodes.

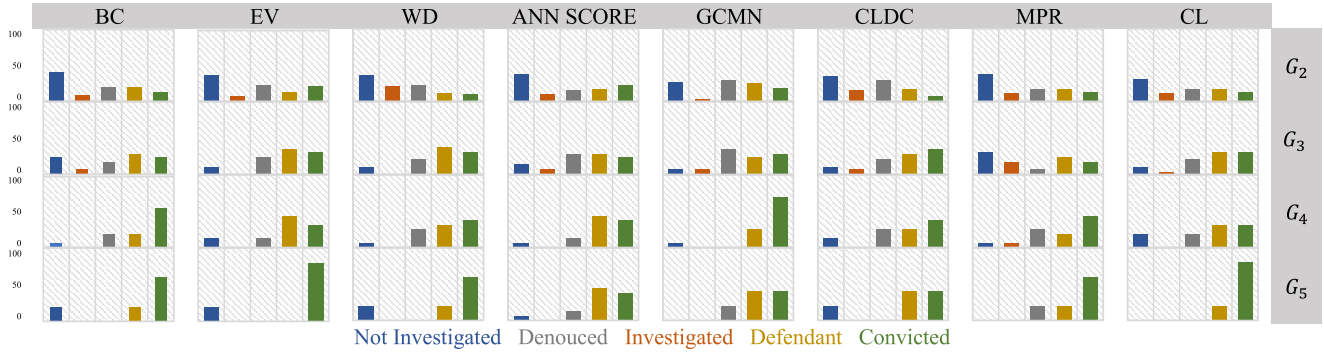



**FIGURE 4.** Comparative analysis of number of nodes per group/legal status.

**TABLE 4.** Relevance numeric analysis between groups.

| Group | Relevance Criterion | BC | EV | WD | ANN SCORE | GCMN | CLDC | MPR | CL |
|---|---|---|---|---|---|---|---|---|---|
| $G_2$ | $MRS$ | 32 | 35 | 27 | 33 | 44 | 32 | 28 | 28 |
| | $LRS$ | 30 | 27 | 35 | 29 | 18 | 30 | 30 | 26 |
| $G_3$ | $MRS$ | 20 | 26 | 26 | 23 | 25 | 24 | 14 | 24 |
| | $LRS$ | 9 | 3 | 3 | 6 | 4 | 5 | 14 | 4 |
| $G_4$ | $MRS$ | 15 | 14 | 15 | 15 | 15 | 14 | 14 | 13 |
| | $LRS$ | 1 | 2 | 1 | 1 | 1 | 2 | 2 | 3 |
| $G_5$ | $MRS$ | 4 | 4 | 4 | 4 | 5 | 4 | 5 | 5 |
| | $LRS$ | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 |

**TABLE 5.** Relevance index (Equation 8) and general relevance (Equation 9) analysis.

| Group | BC | EV | WD | ANN SCORE | GCMN | CLDC | MPR | CL |
|---|---|---|---|---|---|---|---|---|
| $G_2$ | 52 | 56 | 44 | 53 | 71 | 52 | 48 | 52 |
| $G_3$ | 69 | 90 | 90 | 79 | 86 | 83 | 50 | 86 |
| $G_4$ | 94 | 88 | 94 | 94 | 94 | 88 | 88 | 81 |
| $G_5$ | 80 | 80 | 80 | 80 | 100 | 80 | 100 | 100 |
| General Relevance | 63 | 71 | 64 | 67 | 79 | 66 | 57 | 68 |

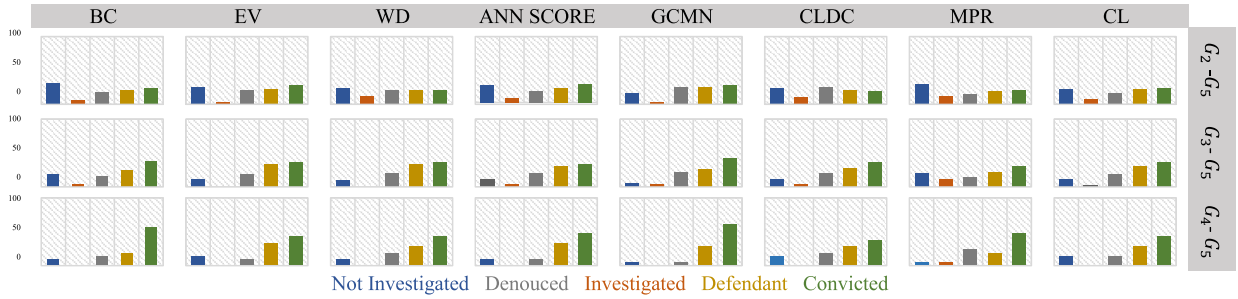



**FIGURE 5.** Comparative cumulative analysis of number of nodes per group/legal status.

This differentiation comes from the use of a completely different strategy from those used by centrality measures like betweenness centrality (BC), eigenvector (EV), closeness (CL), and weighted degree (WD), which at some point have premises based on familiar concepts [4]–[8]. As expected, the CLDC and MPR centrality achieved a better result than

**TABLE 6.** Individuals distribution per legal status and centrality measures.

| Legal Status | BC | EV | WD | ANN SCORE | GCMN | CLDC | MPR | CL |
|---|---|---|---|---|---|---|---|---|
| Convicted | 27 | 31 | 24 | 31 | 32 | 22 | 23 | 26 |
| Defendant | 24 | 25 | 24 | 25 | 28 | 24 | 21 | 25 |
| Denounced | 20 | 23 | 24 | 19 | 29 | 28 | 17 | 19 |
| Total | 71 | 79 | 72 | 75 | 89 | 74 | 61 | 70 |
| Investigated | 7 | 4 | 13 | 8 | 4 | 11 | 13 | 8 |
| Not Investigated / Acquitted | 34 | 29 | 27 | 29 | 19 | 27 | 33 | 25 |
| Total | 41 | 33 | 40 | 37 | 23 | 38 | 46 | 33 |

**TABLE 7.** Qualitative analysis of groups $G_3$ to $G_5$ (*MRS*).

| | BC | EV | WD | GCMN | CLDC | MPR | CL |
|---|---|---|---|---|---|---|---|
| $LRS$ | 6 | 1 | 1 | 4 | 4 | 3 | 1 |
| $MRS$ | 4 | 3 | 4 | 13 | 8 | 8 | 3 |
| Total | 10 | 4 | 5 | 17 | 12 | 11 | 4 |

the standard centralities once their ranking was based on a multi-layered strategy.

It is essential to clarify that the absence of new nodes in the ANN SCORE is the expected result once it represents the normalized geometric mean of the other three metrics (BC), eigenvector (EV), and weighted degree (WD) [19].

## VII. CONCLUSION AND FUTURE WORK

This work proposes the GCMN centrality, an approach centered in groups, to find relevant nodes in scenarios described by a multiplex network. We did the mathematical formalization of the novel centrality using: tensorial notation, algebra of higher order, and regular expressions (sections III-B, IV); its algorithm complexity is $O(n)$, demonstrating the implementation viability (subsection IV-A).

As proof of concept, we used a case study involving five testimonies of prisoners condemned by the most massive anti-fraud operation in Brazil's history, called the Car Wash Operation. According to already established centrality metrics, the modeling of these statements into a complex network resulted in a ranking of most relevant nodes. The GCMN centrality proved to be superior to well-known standard centrality measures: weighted degree (WD) [8], betweenness (BW) [4], [5], closeness (CL) [4], and eigenvector (EV) [6]; and two multilayered centrality measure: Cross-layer degree centrality (CLDC) [21] and The Multiplex PageRank (MPR) [22]; in detecting denounced, defendant or convicted individuals. This analysis was done in a segmented way by groups of nodes — $G_2$ to $G_5$ —, in which the GCMN centrality showed superior results:

- The GCMN centrality ranking achieved 90% of Precision, and 79% of Accuracy; in detecting individuals with the most relevant legal status (*MRS*). The other centrality metrics achieved inferior results in this analysis (Sub-Section VI-A, Fig. 2);
- We used the weight (Equation 2) of a node as a criterion to distribute the nodes into groups (Equation 3). The analysis in Sub-Section VI-B showed that this was the right choice once the degree of severity associated with the legal status of individuals, distributed into the groups, had consistent growth (from $G_2$ to $G_5$) for all the evaluated centrality measures (Fig. 3);
- The GCMN centrality found more novelty's them all the other centralities together. That means that the GCMN reached more significant nodes not pointed by any other centrality (Sub-Section VI-E).
- Finally, we had the analysis of the relevance per group individually and cumulatively (Figs. 4 and 5) ways. Regarding the most important groups ($G_3$ to $G_5$) of individuals and their legal status, the GCMN centrality could point to the most relevant individuals. The GCMN centrality achieved a general best performance in the cumulative analysis, both for the main groups $G_3$ to $G_5$ and $G_4$ to $G_5$, pointing the individuals with the most severe legal status (Sub-Sections VI-C and VI-D).

As following steps, we want to apply the GCMN centrality to other case studies to attest its effectiveness and compare the centrality to other group-based centrality measures [29]–[31], [33].

The database, scripts, analyzes, graphs, and complex networks that supported this work's preparation are available at: https://data.mendeley.com/datasets/28xd6jz46j/draft?a=45e496ce-b8a8-4601-9ffa-c487526a6327.

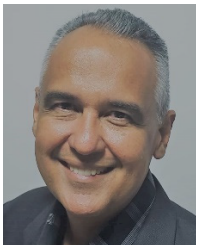
## ACKNOWLEDGMENT

The authors would like to thank Pedro Figueirêdo and Edson Lima Jr. for revising our text and for helping with the mathematical formalization.

## REFERENCES

[1] G Caldarelli, "Scale-free networks: Complex webs," *Nature and Technology*. New York, NY, USA: Oxford Univ. Press, 2020.
[2] K. Das, S. Samanta, and M. Pal, "Study on centrality measures in social networks: A survey," *Social Netw. Anal. Mining*, vol. 8, no. 1, pp. 1–11, Dec. 2018.

[3] A Bavelas, "Communication patterns in task-oriented groups," *J. Acoust. Soc. Amer.*, vol. 22, no. 6, pp. 725–730, 1950.
[4] L. C. Freeman, "Centrality in social networks conceptual clarification," *Social Netw.*, vol. 1, no. 3, pp. 215–239, Jan. 1978.
[5] E. Otte and R. Rousseau, "Social network analysis: A powerful strategy, also for the information sciences," *J. Inf. Sci.*, vol. 28, no. 6, pp. 441–453, Dec. 2002.
[6] P Bonacich, "Factoring and weighting approaches to status scores and clique identification," *J. Math. Sociol.*, vol. 2, no. 1, pp. 113–120, Jan. 1972.
[7] P. Bonacich, "Some unique properties of eigenvector centrality," *Social Netw.*, vol. 29, no. 4, pp. 555–564, Oct. 2007.
[8] A. Beveridge and J. Shan, "Network of thrones," *Math Horizons*, vol. 23, no. 4, pp. 18–22, Apr. 2016.
[9] U Brandes and T Erlebach, *Network Analysis—Methodological Foundations—Introduction*. Berlin, Germany: Springer-Verlag, 2005, doi: 10.1007/b106453.
[10] H. Liao, M. S. Mariani, M. Medo, Y.-C. Zhang, and M.-Y. Zhou, "Ranking in evolving complex networks," *Phys. Rep.*, vol. 689, pp. 1–54, May 2017.
[11] N. A. Christakis and J. H. Fowler, "Social network sensors for early detection of contagious outbreaks," *PLoS ONE*, vol. 5, no. 9, Sep. 2010, Art. no. e12948.
[12] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, "Epidemic processes in complex networks," *Rev. Modern Phys.*, vol. 87, no. 3, pp. 925–979, Aug. 2015.
[13] R. Guimera, S. Mossa, A. Turtschi, and L. A. N. Amaral, "The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles," *Proc. Nat. Acad. Sci. USA*, vol. 102, no. 22, pp. 7794–7799, May 2005.
[14] F Schweitzer, G Fagiolo, D Sornette, F Vega-Redondo, A Vespignani, and D R White, "Economic networks: The new challenges," *Science*, vol. 325, no. 5939, p. 80, 2009.
[15] E. Bullmore and O. Sporns, "Complex brain networks: Graph theoretical analysis of structural and functional systems," *Nature Rev. Neurosci.*, vol. 10, no. 3, pp. 186–198, Mar. 2009.
[16] A. Rinaldo, J. R. Banavar, and A. Maritan, "Trees, networks, and hydrology," *Water Resour. Res.*, vol. 42, no. 6, pp. 1–19, Jun. 2006.
[17] K. Juszczyszyn and G. Kolaczek, "Complex networks monitoring and security and fraud detection for enterprises," in *Proc. IEEE 28th Int. Conf. Enabling Technol., Infrastruct. Collaborative Enterprises (WETICE)*, Jun. 2019, pp. 124–125.
[18] C. Sciarra, G. Chiarotti, F. Laio, and L. Ridolfi, "A change of perspective in network centrality," *Sci. Rep.*, vol. 8, no. 1, Dec. 2018, doi: 10.1038/s41598-018-33336-8.
[19] T. Almeida, F. G. Nakamura, and E. F. Nakamura, "Uma abordagem baseada em redes complexas para análise depoimentos legais," in *Proc. 37th Congresso Sociedade Brasileira Computação*, 2017, pp. 2482–2491.
[20] G Bianconi, *Multilayer Networks: Structure and Function*. Oxford, U.K.: Oxford Univ. Press, 2018, [Online]. Available: https://books.google.com.br/books?id=6v5cDwAAQBAJ
[21] P. Bródka, K. Skibicki, P. Kazienko, and K. Musiał, "A degree centrality in multi-layered social network," 2012, *arXiv:1210.5184*. [Online]. Available: https://arxiv.org/abs/1210.5184
[22] X. Tu, G.-P. Jiang, Y. Song, and X. Zhang, "Novel multiplex PageRank in multilayer networks," *IEEE Access*, vol. 6, pp. 12530–12538, 2018.
[23] A. Solé-Ribalta, M. De Domenico, S. Gómez, and A. Arenas, "Random walk centrality in interconnected multilayer networks," *Phys. D, Nonlinear Phenomena*, vol. 323, pp. 73–79, Jun. 2016, [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167278916000026
[24] M. De Domenico, A. Solé-Ribalta, E. Cozzo, M. Kivelä, Y. Moreno, M. A. Porter, S. Gómez, and A. Arenas, "Mathematical formulation of multilayer networks," *Phys. Rev. X*, vol. 3, no. 4, Dec. 2013, Art. no. 041022, doi: 10.1103/PhysRevX.3.041022.
[25] P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, and J.-P. Onnela, "Community structure in time-dependent, multiscale, and multiplex networks," *Science*, vol. 328, no. 5980, pp. 876–878, May 2010. [Online]. Available: https://www.sciencemag.org/cgi/content/full/sci;328/5980/876/DC2
[26] V. Nicosia, G. Bianconi, V. Latora, and M. Barthelemy, "Growing multiplex networks," 2013, *arXiv:1302.7126*. [Online]. Available: https://arxiv.org/abs/1302.7126
[27] M. De Domenico, A. Sole-Ribalta, S. Gomez, and A. Arenas, "Navigability of interconnected networks under random failures," *Proc. Nat. Acad. Sci. USA*, vol. 111, no. 23, pp. 8351–8356, May 2014, doi: 10.1073/pnas.1318469111.
[28] M. Nanni, L. Tortosa, J. F. Vicent, and G. Yeghikyan, "Ranking places in attributed temporal urban mobility networks," *PLoS ONE*, vol. 15, no. 10, Oct. 2020, Art. no. e0239319.
[29] H. Yang, J.-D. Luo, Y. Fan, and L. Zhu, "Using weighted K-means to identify Chinese leading venture capital firms incorporating with centrality measures," *Inf. Process. Manage.*, vol. 57, no. 2, Mar. 2020, Art. no. 102083.
[30] M.-H. Hsieh and C. L. Magee, "A new method for finding hierarchical subgroups from networks," *Social Netw.*, vol. 32, no. 3, pp. 234–244, Jul. 2010.
[31] S. P. Borgatti and M. G. Everett, "Models of core/periphery structures," *Social Netw.*, vol. 21, no. 4, pp. 375–395, Oct. 2000.
[32] F. Battiston, V. Nicosia, and V. Latora, "Structural measures for multiplex networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 89, no. 3, Mar. 2014, doi: 10.1103/PhysRevE.89.032804.
[33] F. Agneessens, S. P. Borgatti, and M. G. Everett, "Geodesic based centrality: Unifying the local and the global," *Social Netw.*, vol. 49, pp. 12–26, May 2017.



**BRUNO CÉSAR BARRETO DE FIGUEIRÊDO** received the B.Sc. and M.S. degrees in informatics from the Federal University of Paraíba, in 1991 and 2008, respectively. He is currently pursuing the Ph.D. degree in informatics with the Federal University of Amazonas (UFAM). He was a Researcher with Texas A&M University. He is currently with the Roraima Court of Accounts, Brazil, and also with the State University of Roraima, Brazil. His research interests include audit of systems and public accounts, complex networks, and distributed systems.

**FABÍOLA GUERRA NAKAMURA** received the degree in engenharia elétrica from the Universidade Federal do Amazonas, in 1997, the master's degree in ciência da computação and the Ph.D. degree in doutorado em ciência da computação from the Universidade Federal de Minas Gerais, in 2003 and 2010, respectively. She has experience in computer science. She is currently with the Institute of Computing, Federal University of Amazonas, Manaus, Brazil.

**EDUARDO FREIRE NAKAMURA** received the M.S. and Ph.D. degrees in computer science from the Federal University of Minas Gerais (UFMG). He was a Visiting Associate Professor with Texas A&M University. He is currently a Professor of computer science and engineering with the Federal University of Amazonas (UFAM). He is also with the Institute of Computing, Federal University of Amazonas, Manaus, Brazil. His research interests include data science, network science, and wireless sensor networks. He was awarded as the best Ph.D. thesis in engineering and exact and earth sciences.