Article

Combining the Two-Layers PageRank Approach with the APA Centrality in Networks with Data

Taras Agryzkov [1], Francisco Pedroche [2], Leandro Tortosa [1] and José F. Vicent [1,*]

[1] Department of Computer Science and Artificial Intelligence, Campus de San Vicente, University of Alicante, Ap. Correos, 99 E-03080 Alicante, Spain; taras.agryzkov@ua.es (T.A.); tortosa@ua.es (L.T.)

[2] Institut de Matemàtica Multidisciplinària, Universitat Politècnica de València, Camino de Vera, s/n, E-46022 València, Spain; francisco.pedroche@upv.es

* Correspondence: jvicent@ua.es; Tel.: +34-965903900

Abstract: Identifying the influential nodes in complex networks is a fundamental and practical topic at the moment. In this paper, a new centrality measure for complex networks is proposed based on two contrasting models that have their common origin in the well-known PageRank centrality. On the one hand, the essence of the model proposed is taken from the Adapted PageRank Algorithm (APA) centrality, whose main characteristic is that constitutes a measure to establish a ranking of nodes considering the importance of some dataset associated to the network. On the other hand, a technique known as two-layers PageRank approach is applied to this model. This technique focuses on the idea that the PageRank centrality can be understood as a two-layer network, the topological and teleportation layers, respectively. The main point of the proposed centrality is that it combines the APA centrality with the idea of two-layers; however, the difference now is that the teleportation layer is replaced by a layer that collects the data present in the network. This combination gives rise to a new algorithm for ranking the nodes according to their importance. Subsequently, the coherence of the new measure is demonstrated by calculating the correlation and the quantitative differences of both centralities (APA and the new centrality). A detailed study of the differences of both centralities, taking different types of networks, is performed. A real urban network with data randomly generated is evaluated as well as the well-known Zachary's karate club network. Some numerical results are carried out by varying the values of the α parameter—known as *dumping factor* in PageRank model—that varies the importance given to the two layers (topology and data) within the computation of the new centrality. The proposed algorithm takes the best characteristics of the models on which it is based: on the one hand, it is a measure of centrality, in complex networks with data, whose calculation is stable numerically and, on the other hand, it is able to separate the topological properties of the network and the influence of the data.

Keywords: networks centrality; Adapted PageRank Algorithm; PageRank; two-layers PageRank; spectral theory

1. Introduction

1.1. Motivation

How to effectively identify influential nodes (or edges) in complex networks has been paid great attention because of great theoretical significance [1]. The idea of importance of a vertex in complex networks is associated with the mathematical concept of centrality. It constitutes one of the fundamental questions in characterizing the structure and dynamics of complex networks. A high percentage of the existing centrality measures base their criteria of importance of a node on different aspects related to topological properties of the network. The influence of network's data in the

calculation of centrality is not considered in most of them. For instance, let us think about the urban network of a city and its topological representation by means of a primal graph [2]; we realize that the data associated with the city, whether real or virtual, define an important part of its essence and they must not be omitted. Therefore, it is necessary to dispose of efficient centrality measures that take into account both the connectivity of the nodes and the information associated with them.

1.2. Literature Review

The first centrality measures were developed in social networks [3], many of them were introduced to reflect their sociological origin [4,5]. Nowadays, they have become an important tool in network analysis, and are widely used for ranking the influence in social networks [6,7], Internet web-page popularity [8], computer networks [9], spread of epidemic diseases [10], ranking reputation of scientists [11], urban networks [12–14], etc.

Depending on the specific application, different centrality measures may be of interest. There are several studies in the literature that aim at defining widely used network centrality measures based on network graphs. For instance: degree centrality [15], eigenvector centrality [16] closeness centrality [17], betweenness centrality [18,19].

Degree centrality is defined as the number of edges incident upon a vertex. As advantages it simplicity and low computing complexity. However, degree centrality has some limitations, such as: the measure does not take the global structure of the graph into consideration. Similar advantages and disadvantages have the eigenvector centrality. It is a natural extension of degree centrality and it is also a measure of the influence of a node in a network. A high eigenvector score means that a node is connected to many nodes who themselves have high scores. Closeness centrality is defined as the inverse of the sum of shortest distances to all other vertices from a focal vertex. It can be treated as a measure of how efficiently it exchanges information with others in a graph. The main limitation is the lack of applicability to graphs with disconnected components. Betweenness centrality is the number of all shortest paths that pass through a vertex. It was introduced by Freeman [18] as a measure for quantifying the control of a human on the communication among other humans in a social network and its main weakness is the difficulty of being applied in large scale networks.

One of the centrality measures proposed two decades ago is the well-known PageRank centrality. The *PageRank* method [8] was proposed to compute a ranking for every Web page based on the Web's graph. It constitutes a global ranking of all Web pages, regardless of their content, based exclusively on their location in the Web's graph structure. The purpose of the method is obtaining a vector, called *PageRank vector*, which gives the relative importance of the pages. Since this vector is calculated based on the structure of the Web connections, it is independent of the request of the person performing the search.

Over the years, some modifications of this model have been proposed. For instance, in [20] the authors develop a more accurate search results computing a set of PageRank vectors. In [21], a link-based algorithm is built on a random surfer model reflecting back steps made by a Web. To further details regarding the theoretical foundations of the PageRank formulation, see [22–24].

In [25], the authors disclose the fundamental properties concerning stability, complexity of computational scheme, and critical role of parameters involved in the computationof the PageRank. This algorithm uses the *Power method* to compute successive iterates that converge to the principal eigenvector of the Markov chain representing the Web link graph. In [26], the authors present an effective heuristic relaxed and extrapolated algorithm based on the Power method that accelerates the convergence.

In recent years, new studies have been developed on the subject of centrality in various types of complex networks, as well as networks with data [27–29]. In [27], the authors propose a new centrality measure (called Adapted PageRank Algorithm -APA-) which main contribution is to establish a ranking of nodes, from the point of view of the topology and the data associated to the nodes. This model

shares the concept of PageRank vector, taking into account not only the connectivity between different network nodes, but also external characteristics defined for each specific problem.

Many real-life systems can be modelled by taking into account the fact that the interrelations between nodes are heterogeneous [30,31]. This heterogeneity makes that some structural and dynamical properties emerge from the distinction between different kinds of links, and a new model of multilayer is introduced [32,33]. With the aim to define a new centrality in that type of multiplex networks, a new approach to the PageRank analysis was developed [34]. Throughout this process, the authors consider that the PageRank vector can be understood as the stationary distribution of a Markov chain that occurs in a network with two layers: the physical layer (the network) and the teleportation layer.

1.3. Main Contribution

The idea of two-layers may be combined with the APA centrality model, where it is reasonable to consider the centrality understood under the perspective of two layers: the topology layer defined by the network and the data layer defined by the data associated to the network.

In this paper, the procedure for developing a new centrality measure denoted as APA2f is introduced. This procedure is based on the combination of two known methods. On the one hand, the idea of APA centrality is used and, on the other hand, this centrality is adapted to the concept of two-layers approach. The proposed centrality has the advantage of combining the best properties of both models: it is able to establish a ranking of nodes in complex networks with data and interprets the classic PageRank centrality as a multilayer network model. The design and analysis of APA2f measure constitutes the main objective of this work.

1.4. Structure of the Paper

This paper is organized as follows: Section 2 summarizes the basic characteristics of the APA centrality, as well as the two-layers approach to the PageRank vector concept. Once these well-known models are described, the new centrality is presented by combining the idea of centrality APA based on data and the idea of understanding the PageRank vector by the two-layers approach. In Section 3, a concise and precise description of the obtained results, their interpretation and some experimental conclusions are provided. The discussion of the experimental results as well as a deep study of the differences in both centralities for networks with different sizes are performed in Section 4. Finally, some conclusions are summarised.

2. Materials and Methods

In this section, both the well-established methods and the proposed new centrality method for networks with data are described in detail. The model presented has the main characteristic that can be understood as the merger of two well-established and consolidated models for the calculation of centrality. On the one hand, it takes as a base the APA centrality for networks with data. The great advantage of this model is that it is based on the idea of the PageRank vector, whose essence is the calculation of the dominant eigenvector of an irreducible and stochastic matrix. By choosing this matrix conveniently, its stability, from the numerical point of view, is absolute and there exist efficient numerical methods that obtain results although the networks are very large. On the other hand, the idea of considering the PageRank model as the union of two different layers, such as topological and probabilistic, leads us directly to our approach of incorporating a layer of data in the network as essential when determining the importance of nodes in the same.

Throughout this section, it is described the two approaches on which the new centrality is based and, later, an algorithm to calculate the new measure will be implemented, exposing its main characteristics.

2.1. The Original APA Model Based on Data

The Adapted PageRank Algorithm (APA) proposed by Agryzkov et al. [27] provides us a model to establish a ranking of nodes in an urban network taking into account the data presented in it. This centrality was originally proposed for urban networks, although it may be generalized to spatial networks or networks with data. It constitutes a centrality measure for networks with the main characteristic that it is able to consider the importance of data obtained from any source in the whole process of computing the centrality of the individual nodes. Starting from the basic idea of the PageRank vector concept, the construction of the matrix used for obtaining the classification of the nodes is modified.

In its original approach, PageRank is based on a model of a Web surfer that probabilistically browses the Web graph, starting at a node chosen at random according to a *personalization vector* whose components give us the probability of starting at node v. At each step, if the current node has outgoing links to other nodes, the surfer next browses with probability α one of those nodes (chosen uniformly at random), and with probability $1 - \alpha$ a node chosen at random according to the personalized vector. For the Web graph, the most popular value of the dumping factor is 0.85. If the current node is a sink with no outgoing links, the surfer automatically chooses the next node at random according to the personalized vector.

In the APA model, the data matrix is constructed following a similar reasoning from the original idea of the PageRank vector; a random walker can jump between connecting nodes following the local link given by the network or can jump between nodes (not directly connected) with the same probability, regardless the topological distance between them (number of nodes in the walk).

In the algorithm implemented to calculate the APA centrality (see [27], page 2190), a new matrix $A^* = \left(p_{ij} \right) \in \mathcal{R}^{n \times n}$ is constructed from the adjacency matrix A, as

$$p_{ij} = \begin{cases} \frac{1}{c_j} & \text{if } a_{ij} \neq 0, \\ 0 & \text{otherwise,} \end{cases} \quad 1 \leq i, j \leq n, \tag{1}$$

where c_j represents the sum of the j-th column of the adjacency matrix.

Algebraically, A^* may be obtained as

$$A^* = A \Delta^{-1}, \tag{2}$$

where $\Delta = \left(\delta_{ij} \right) \in \mathcal{R}^{n \times n}$ is the degree matrix of the graph, that is, $\delta_{ij} = c_j^{-1}$, for $i = j$ and $\delta_{ij} = 0$, for $i \neq j$. We refer to A^* as *transition matrix*, and it represents, by columns, the probability to navigate from a page to other. In the literature related to this topic the matrix A^* is also denoted as P or P_A, so A^*, P or P_A are the same matrix. Following the notation of Pedroche et al. it will preferably be used P.

The transition matrix, $P = A^*$ has the following characteristics (see [27]):

1. It is *nonnegative*.
2. It is *stochastic* by columns.
3. The highest eigenvalue of P is $\lambda = 1$.

The key point of the model is the construction of the so-called *data matrix D* of size $n \times k$, with its n rows representing the n nodes of the network, and each of its k columns representing the attributes of the data object of the study. Specifically, an element $d_{ij} \in D$ is the value we attach to the data class k_j at node i.

However, not all the characteristics of data may have the same relevance or influence in the question object of the analysis. Therefore, a vector $\vec{v_0} \in \mathbb{R}^{k \times 1}$ is constructed, where the element that occupies the row i is the multiplicative factor associated with the property or characteristic k_i. With this vector $\vec{v_0}$ a weighting factor of the data is introduced, in order to work with the entire data set or a part of it.

Then, multiplying D and $\vec{v}_0$, may be obtained $\vec{v}$ as

$$\vec{v} = D \cdot \vec{v}_0,$$

with $\vec{v} \in \mathbb{R}^{n \times 1}$.

The construction of vector $\vec{v}$ allows us to associate to each node a value that represents the amount of data assigned to it. Thus, two different values are associated with every node; on the one hand, its degree, related to the topology and, on the other hand, the value of the data associated to it. For a more detailed description of how the data are associated to the nodes, see [27,29].

After normalizing $\vec{v}$, denoted as $\vec{v}^*$, it is possible to define the matrix M_{APA} as

$$M_{APA} = (1 - \alpha)P + \alpha V, \tag{3}$$

where $V \in \mathbb{R}^{n \times n}$ is a matrix in which all of its components in the i-th row are equal to $\vec{v}_i^*$. The parameter α is fixed and it is related to the teleportation idea. The value that is traditionally used is $\alpha = 0.15$.

In practice, vector $\vec{v}^*$ is repeated (n times) in every column of the matrix V.

The matrix M_{APA} is used to compute the ranking vector for the network.

With these considerations, the APA algorithm proposed in [27] may be summarized as:

Algorithm 1: APA

Let $G = (V, E)$ be a primary graph representing a network with n nodes.

1. Compute the matrix P from the graph G.
2. Construct the data matrix D.
3. Construct the weighted vector $\vec{v}_0$.
4. Compute $\vec{v}$ as $D\vec{v}_0 = \vec{v}$.
5. Normalize $\vec{v}$, and denote it as $\vec{v}^*$.
6. Construct V as $V = \vec{v}^* \vec{e}^T$.
7. Construct the matrix M_{APA} as $M_{APA} = (1 - \alpha)P + \alpha V$.
8. Compute the eigenvector $\vec{x}$ of the matrix M_{APA} associated to eigenvalue $\lambda = 1$. The components of the resulting eigenvector $\vec{x}$ represent the ranking of the nodes in the graph G.

The main feature of this algorithm is the construction of the data matrix D and the weighted vector $\vec{v}_0$. The matrix D allows us to represent numerically the dataset. Vector $\vec{v}_0$ determines the importance of each of the factors or characteristics that have been measured by means of D.

The Perron-Frobenius theorem is of great importance in this problem, since it constitutes the theoretical base that ensures that there exists an eigenvector $\vec{x}$ associated with the dominant eigenvalue $\lambda = 1$, so that all its components are positive, which allows establishing an order or classification of these elements. In our case, due to the way in which P and V have been constructed, it can be seen that M_{APA} is a stochastic matrix by columns, which assures us of the spectral properties necessary for the Perron-Frobenius theorem to be fulfilled. Therefore, the existence and uniqueness of a dominant eigenvector with all its components positive is guaranteed. See [8,35] for further study of spectral and algebraic properties of the models based on PageRank.

Vector $\vec{x}$ constitutes the Adapted PageRank vector and provides a classification or ranking of the pages according to the connectivity criterion between them and the presence of data.
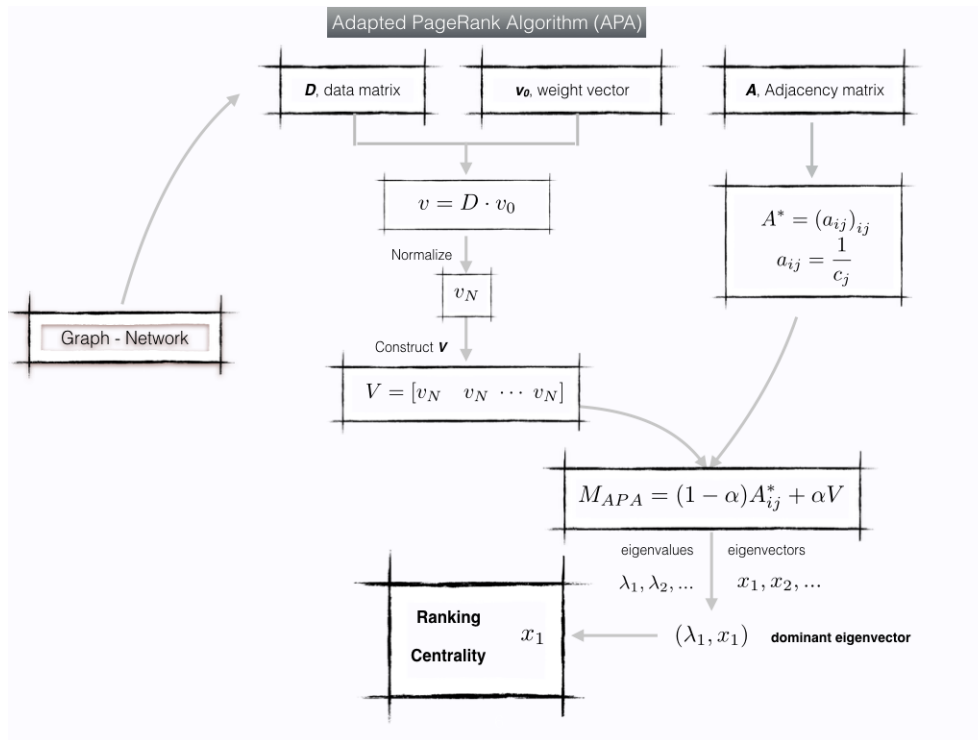
We can summarize this model in Figure 1.

Figure 1. Scheme of the APA method.

2.2. The Two-Layers Approach for Classic PageRank

In practical applications, may be interesting to assign a global measure of importance to each node. If the system contains several types of relations between actors it is expected that the measures, in some way, consider the importance obtained from the different layers. A simple choice could be to combine the centrality of the nodes, obtained from the different layers independently according to some heuristic choice.

Pedroche et al. [34] proposed a two-layers approach for the classic PageRank classification vector based on the idea that will be explained in this section. The key is to consider the PageRank model as a process divided into two clearly differentiated parts. In the first place, there is a part related to the topology of the network, where the connections of the nodes are basically taken into account by means of their adjacency matrix. Secondly, there is a part regarding to the probability for jumping between two nodes in the network, following a criterion that there is the same probability among all of them.

In [34], the authors realize that the PageRank classification for a graph G with personalized vector $\vec{v}$ can be understood as the stationary distribution of a Markov chain that occurs in a network with two layers, which are

l_1, **a physical layer:** the network G.
l_2, **a teleportation layer:** an all-to-all weighted network given by the personalized vector.

Under this perspective, it is easy to construct a block matrix M_A based on these two-layers where each of the diagonal blocks is associated to a given layer. Therefore, M_A can be constructed as

$$M_A = \left(\begin{array}{c|c} \alpha P_A & (1-\alpha)I \\ \hline \alpha I & (1-\alpha)\vec{e}\vec{v}^T \end{array} \right) \quad \in \mathbb{R}^{2n \times 2n}. \tag{4}$$

where M_A defines a Markov chain in a network with two layers.

Because of the good spectral characteristics of M_A (it is irreducible and primitive), Pedroche et al. [34] arrive to the conclusion that given a network with n nodes and whose adjacency matrix is A, the two-layer approach PageRank of A is the vector

$$\hat{\pi}_A = \pi_u + \pi_d \in \mathbb{R}^n,$$

where $\begin{bmatrix} \pi_u^T & \pi_d^T \end{bmatrix}^T \in \mathbb{R}^{2n}$ is the unique normalized and positive eigenvector of matrix M_A given by the expression (4).

The idea of separating the centrality based on the PageRank concept into two layers, differentiating the topological part of the network from the concept of personalization vector, can be extrapolated to multilayer networks, as the authors demonstrate in [34].

This idea motivates the adaptation of the APA algorithm to the consideration of the division of the process in two different layers, where now the layer related to the topology is not modified, while the layer defined by the personalization vector is replaced by the layer of data in the network. In the following subsections, this idea will be developed and analysed both theoretically and numerically.

2.3. Constructing the APA Centrality by Applying the Two-Layer Technique

Roughly speaking, the idea in which the PageRank classification vector is based in that of a vector of a random walker that can jump between nodes following two basic principles: on the one hand, the walker can use the local links (physical walk) and, on the other hand, the walker can jump to any other node in the network according to the personalised vector (teleportation). This motivates the two-layers approach.

The APA centrality model is also based on the PageRank vector concept. However, the idea of a teleportation layer is replaced by the influence of data present in the network and associated to the nodes. This is motivated by the idea of applying this centrality to spatial or urban networks, where the analysis and the amount of data in the network is more important than the concept of random navigation through the network. It is assumed that in a network with data there is a greater probability that a user will move not randomly but following the data, according to their interests. This idea is what fundamentally differentiates Pedroche et al. and APA models.

Although both models are different, the application of the two-layers PageRank approach model as it was described in Section 2.2 in the APA centrality model makes sense, as it is explained below. Analysing carefully the construction of matrix M_{APA} given by Equation (3), it can be observed that the influence of the data and the topology are measured separately. This is the basis on which it will be possible to build a new centrality based on the two-layers concept with two clearly differentiated layers:

l_1, **a physical layer:** or topological layer, the network G.
l_2, **a data layer:** the data present in the network.

Therefore, in this section a new centrality algorithm is described with the primary objective to adapt the two-layers approach to the APA centrality, constructing a block matrix to distinguish in the network both topology and data.

In our case, the basis of the original APA model consists of the construction of a stochastic matrix by columns, where the topology of the network is reflected and the influence of the data, through the matrix V.

Let us develop a similar reasoning as in [34], in order to build a 2×2 block matrix where the first upper diagonal block contains the information referring to the network topology, while the lower diagonal block is associated to the collected data in the network and assigned to each node.

Taking as a reference the APA algorithm, the matrix used to compute the eigenvector associated to the dominant eigenvalue $\lambda = 1$ is given by

$$M_{APA} = (1 - \alpha)P + \alpha V,$$

following the notation that appears in [34] regarding to the probability matrix P.

A new 2×2 block matrix is defined as

$$M_{APA2} = \left(\begin{array}{c|c} (1-\alpha)P & (1-\alpha)I \\ \hline \alpha I & \alpha V \end{array} \right) \quad \in \mathbb{R}^{2n \times 2n}. \tag{5}$$

Note the differences between matrices M_{APA2} y M_A. One would expect both to be the same but this is not the case because in the two-layers approach an stochastic by rows matrix M_A is constructed; however, the new centrality proposed is based on the construction of M_{APA2}, an stochastic by columns matrix.

Due to the way it has been constructed, Equation (5) allows us to confirm that M_{APA2} still retains some desirable spectral properties, since M_{APA2} is again irreducible and stochastic by columns. Applying again the Perron-Frobenius theorem it can be said that there exists an eigenvector $\pi_{APA2} \in \mathbb{R}^{2n}$ such that

$$M_{APA2} \, \pi_{APA2} = \pi_{APA2},$$

with

$$\pi_{APA2} = [\pi_u \, \pi_d]^T \in \mathbb{R}^{2n}.$$

Then, the centrality is given by the unique normalized vector $\vec{x}$ given by

$$\vec{x} = (\pi_u + \pi_d) \in \mathbb{R}^n.$$

The new algorithm proposed to implement the APA centrality following the two-layers approach is given by

---
Algorithm 2: APA2f

Let $G = (V, E)$ be a primal graph representing a network with n nodes.

1   Compute the matrix $P = A^*$ from the graph G.
2   Construct the data matrix D.
3   Construct the weighted vector $\vec{v}_0$.
4   Compute $\vec{v}$ as $D\vec{v}_0 = \vec{v}$.
5   Normalize $\vec{v}$, and denote it as $\vec{v}^*$.
6   Construct V as $V = \vec{v}^* \vec{e}^T$.
7   Construct the matrix M_{APA2} using the expression (5).
8   Compute the eigenvector $\vec{x}$ of matrix M_{APA2} associated to eigenvalue $\lambda = 1$.
      The normalized components of the resulting eigenvector $\vec{x}$ represent the ranking of the nodes in graph G.

Algorithm 2 implements a new centrality measure that will be denoted in the following as APA2f. The idea that underlies the construction of the matrix by blocks given by (5) is to maintain the spectral properties of the original matrix M_{APA}, in order that the numerical algorithms for determining dominant eigenvalue and eigenvector are stable and fast. Note that the new matrix M_{APA2} have doubled the size of the original matrix.

Let us show a simple example to clarify this process, considering a simple graph with 4 nodes whose adjacency matrix is

$$A = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix}$$

and whose data matrix D and weighted vector $\vec{v}_0$ are

$$D = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \vec{v}_0 = [1].$$

From the adjacency matrix it is easy to compute P as

$$P = \begin{bmatrix} 0 & 1/3 & 1 & 0 \\ 0 & 0 & 0 & 1/2 \\ 0 & 1/3 & 0 & 1/2 \\ 1 & 1/3 & 0 & 0 \end{bmatrix}.$$

Now, let us compute V from D and $\vec{v}_0$.

$$\vec{v} = D\vec{v}_0 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix},$$

Let us normalize $\vec{v}$ and construct V as

$$V = \begin{bmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{bmatrix}.$$

Now, from P and V it is possible to construct M_{APA2} by the expression (5).

$$M_{APA2} = \left[\begin{array}{cccc|cccc} 0 & 1/6 & 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/4 & 0 & 1/2 & 0 & 0 \\ 0 & 1/6 & 0 & 1/4 & 0 & 0 & 1/2 & 0 \\ 1/2 & 1/6 & 0 & 0 & 0 & 0 & 0 & 1/2 \\ \hline 1/2 & 0 & 0 & 0 & 1/8 & 1/8 & 1/8 & 1/8 \\ 0 & 1/2 & 0 & 0 & 1/8 & 1/8 & 1/8 & 1/8 \\ 0 & 0 & 1/2 & 0 & 1/8 & 1/8 & 1/8 & 1/8 \\ 0 & 0 & 0 & 1/2 & 1/8 & 1/8 & 1/8 & 1/8 \end{array} \right].$$

Notice that M_{APA2}, given by (5), is a stochastic matrix by columns. Therefore, M_{APA2}, provides us the eigenvector associated to eigenvalue $\lambda = 1$. Once this vector is normalized, the centrality of the nodes is determined. For this example, the result is

$$\vec{x} = (\vec{\pi}_u + \vec{\pi}_d) \in \mathbb{R}^n = [0.2700, 0.2023, 0.2334, 0.2944]^T.$$

3. Results

Some numerical results of the new centrality APA2f calculated running the Algorithm 2 on different networks are presented. All the numerical tests have been carried out by implementing Algorithm 2 in R [36], a Free Software under the terms of GNU project. It constitutes a language and environment especially efficient for computing and graphics.

More specifically, tests are performed on networks of different sizes whose adjacency matrix has been generated randomly. Another test has been performed taking a small real urban network of the city of Murcia, Spain; finally, the well-known network representing the social relations in Zachary's

karate club has been used to obtain the centrality. Note that in the numerical results we also calculate the APA centrality since it is the basis of the new APA2f centrality, which will allow us to carry out a comparison between them.

The networks used for these first numerical results have been generated randomly although they do not represent well-known models of random networks such as Erdős-Rényi. The graph was generated from a function that constructs the adjacency matrix following a random series of 0 and 1, determining the connectivity of the vertices.

To begin, the small network represented in Figure 2 is analysed with detail. It has been randomly generated and it has 25 nodes. At first sight, it is observed that the node with the least connectivity is node 21, which has degree 3. The assignment of the amount of data to each node has also been done in a random way, taking values in a range between 0 and 10, both inclusive.
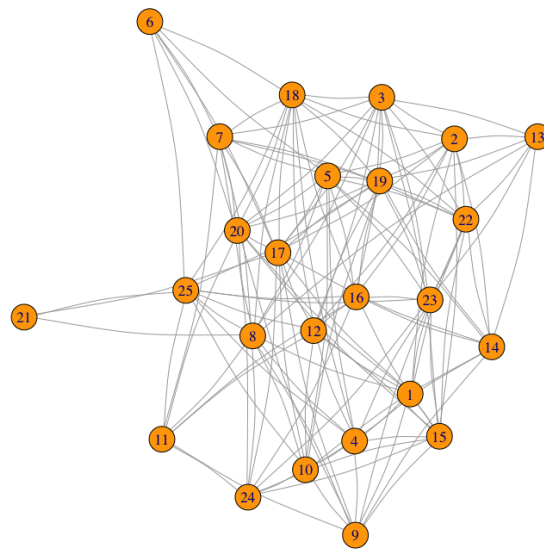
Figure 2. A graph randomly generated with 25 nodes.

The data vector associated to the 25 nodes is

$$D = [6, 4, 2, 5, 10, 10, 0, 7, 8, 9, 5, 0, 9, 2, 4, 10, 1, 6, 0, 10, 7, 6, 9, 4, 10]^T$$

and the weighted vector is chosen equal to 1.

The APA and APA2f centralities are calculated independently, for this particular network with this data configuration given by D. The main reason to calculate both centralities is twofold: on the one hand, to determine the coherence in the results of the classifications provided by the two measures and, on the other hand, to determine the magnitude of differences in the results of both centralities. This will be discussed in the next section.

The numerical study has been developed for different values of the parameter α, which controls the importance that we assign in the calculation of centralities to the topology and data. Remark that the greater the parameter is, the greater importance for data is assumed.

Table 1 summarizes the numerical results for the graph displayed in Figure 2. In the first column, the identifier of the node is shown; in the second and third column, the numerical values of the degree and the amount of data are displayed, respectively. The rest of the columns summarize the results obtained for the APA and APA2f centralities, taking different values of the α parameter, that is, $\alpha = 0.15, 0.3, 0.5, 0.6, 0.85$. The columns labelled as *Diff* represent the difference (in percentages) between the APA and APA2f centrality values. The first three nodes in the classification of both centrality measures are highlighted, (red, blue and green for the first, second and third positions, respectively, in the classification).

In Figures 3 and 4, the APA and APA2f centrality values are displayed, for the 25 nodes of the network, taking different values of the parameter.

Table 1. APA, APA2f and differences between them (in percentage) for the graph of Figure 2.

| n	dg	Data	APA	APA2f	Diff	APA	APA2f	Diff	APA	APA2f	Diff	APA	APA2f	Diff	APA	APA2f	Diff
			$\alpha = 0.15$			$\alpha = 0.3$			$\alpha = 0.5$			$\alpha = 0.6$			$\alpha = 0.85$		
1	10	6	0.0367	0.0366	0.24	0.0372	0.0371	0.15	0.03802	0.0384	0.92	0.0386	0.0392	1.72	0.0403	0.0412	2.22
2	12	4	0.0400	0.0427	6.38	0.0369	0.0397	7.06	0.0335	0.0341	1.92	0.0321	0.0317	1.17	0.0292	0.0282	3.20
3	13	2	0.0414	0.0456	9.21	0.0359	0.0402	10.85	0.0290	0.0295	1.70	0.0257	0.0241	6.33	0.0181	0.0152	15.94
4	12	5	0.0429	0.0435	1.25	0.0418	0.0424	1.36	0.0401	0.0398	0.55	0.0391	0.0383	2.01	0.0364	0.0353	3.18
5	13	10	0.0509	0.0485	4.69	0.0543	0.0518	4.58	0.0588	0.0587	0.17	0.0609	0.0622	2.08	0.0663	0.0684	3.14
6	6	10	0.0292	0.0242	17.07	0.0363	0.0312	14.05	0.0458	0.0458	0.09	0.0504	0.0535	5.69	0.0623	0.0672	7.31
7	11	0	0.0345	0.0383	10.03	0.0286	0.0326	12.30	0.0206	0.0205	0.69	0.0165	0.0139	16.17	0.0062	0.0020	68.21
8	14	7	0.0535	0.0515	3.69	0.0548	0.0526	4.05	0.0551	0.0534	3.14	0.0547	0.0528	3.38	0.0517	0.0496	4.11
9	10	8	0.0392	0.0374	4.56	0.0419	0.0400	4.31	0.0455	0.0457	0.42	0.0474	0.0488	2.85	0.0524	0.0546	3.95
10	11	9	0.0427	0.0411	3.81	0.0454	0.0439	3.33	0.0494	0.0502	1.44	0.0517	0.0538	3.93	0.0580	0.0611	5.01
11	8	5	0.0303	0.0295	2.70	0.0313	0.0305	2.53	0.0324	0.0323	0.24	0.0329	0.0331	0.82	0.0340	0.0345	1.38
12	12	0	0.0365	0.0416	12.21	0.0294	0.0348	15.29	0.0205	0.0210	2.12	0.0162	0.0139	14.43	0.0059	0.0019	68.11
13	7	9	0.0300	0.0271	9.74	0.0349	0.0319	8.56	0.0420	0.0427	1.52	0.0458	0.0488	6.09	0.0559	0.0604	7.43
14	10	2	0.0333	0.0354	6.13	0.0301	0.0322	6.61	0.0258	0.0255	1.14	0.0236	0.0219	7.26	0.0177	0.0151	14.80
15	12	4	0.0412	0.0430	4.20	0.0388	0.0405	4.34	0.0357	0.0356	0.25	0.0342	0.0331	3.22	0.0303	0.0286	5.73
16	12	10	0.0465	0.0448	3.71	0.0497	0.0480	3.45	0.0545	0.0552	1.29	0.0571	0.0594	3.89	0.0645	0.0679	4.99
17	15	1	0.0487	0.0527	7.54	0.0423	0.0463	8.65	0.0332	0.0324	2.47	0.0283	0.0245	13.48	0.0154	0.0096	37.54
18	13	6	0.0462	0.0471	1.96	0.0450	0.0460	2.25	0.0436	0.0440	0.77	0.0431	0.0430	0.08	0.0420	0.0418	0.55
19	15	0	0.0464	0.0521	10.98	0.0381	0.0440	13.33	0.0273	0.0272	0.15	0.0219	0.0184	15.98	0.0083	0.0026	68.22
20	12	10	0.0475	0.0450	5.27	0.0512	0.0487	4.95	0.0562	0.0564	0.32	0.0587	0.0605	2.96	0.0653	0.0681	4.13
21	3	7	0.0169	0.0128	24.28	0.0227	0.0185	18.40	0.0301	0.0301	0.03	0.0337	0.0361	6.61	0.0429	0.0468	8.26
22	12	6	0.0427	0.0435	1.99	0.0419	0.0428	1.97	0.0414	0.0418	0.80	0.0414	0.0416	0.46	0.0415	0.0416	0.35
23	11	9	0.0437	0.0413	5.69	0.0473	0.0446	5.65	0.0520	0.0517	0.50	0.0521	0.0554	2.10	0.0595	0.0616	3.28
24	9	4	0.0320	0.0326	1.84	0.0311	0.0318	2.19	0.0300	0.0301	0.96	0.0294	0.0293	0.54	0.0283	0.0279	1.21
25	11	10	0.0471	0.0420	10.68	0.0530	0.0477	10.09	0.0595	0.0580	2.56	0.0622	0.0626	0.58	0.0674	0.0688	1.97
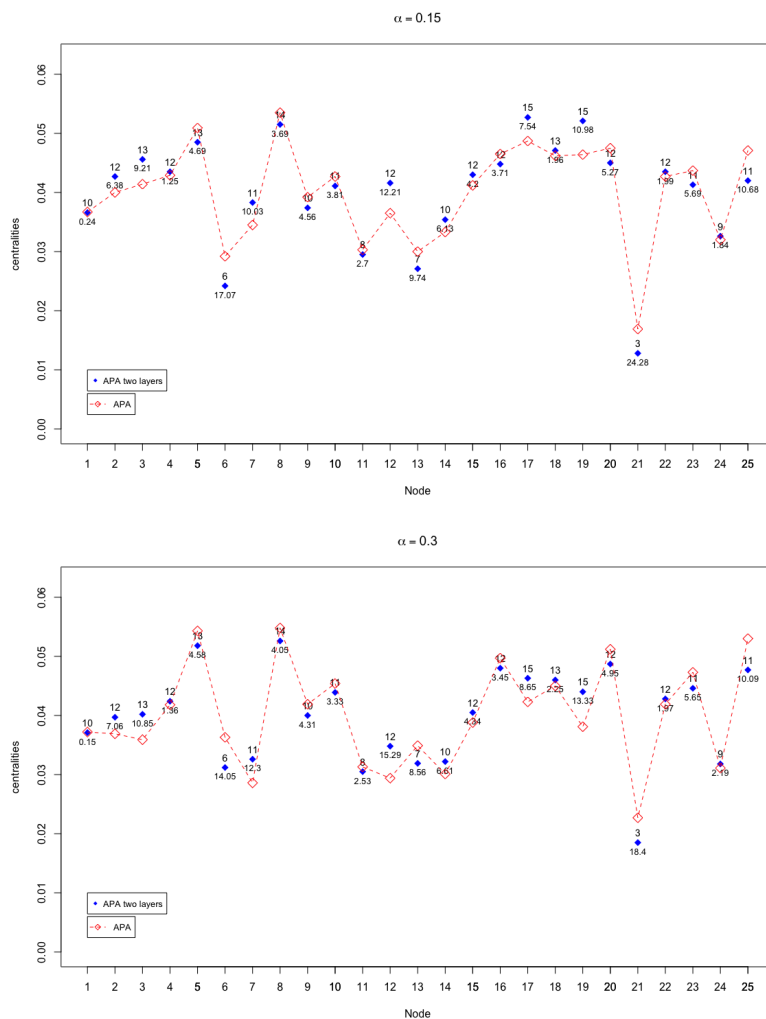
Figure 3. *Cont.*
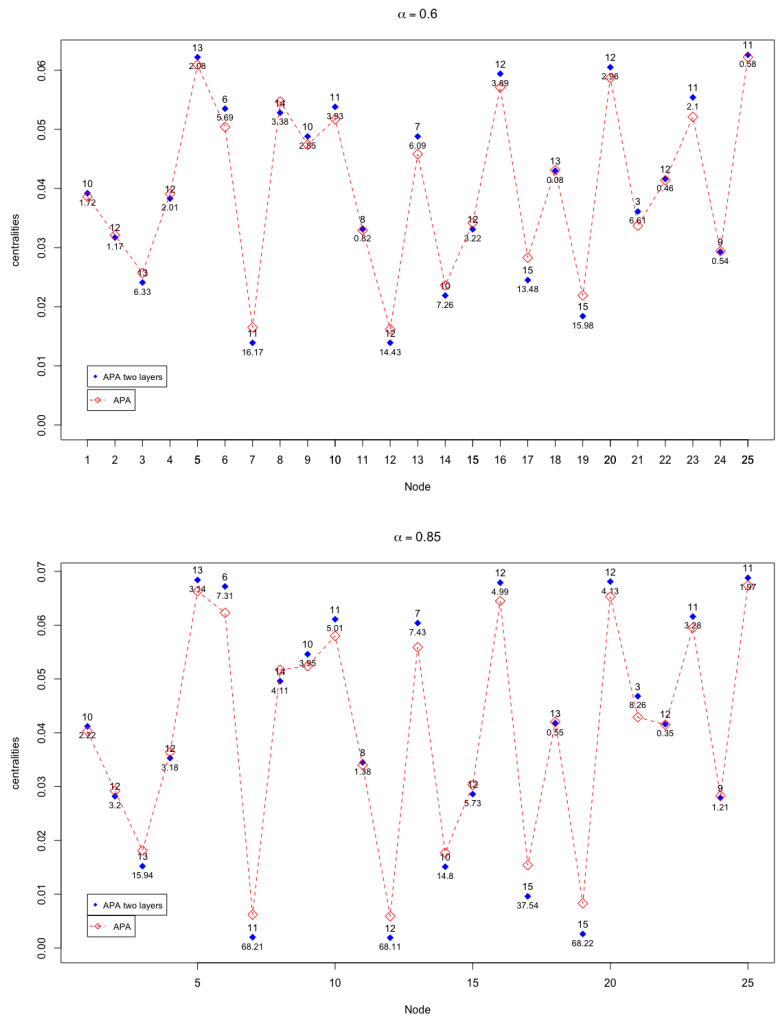
Figure 3. APA and APA2f centralities for the graph represented in Figure 2.
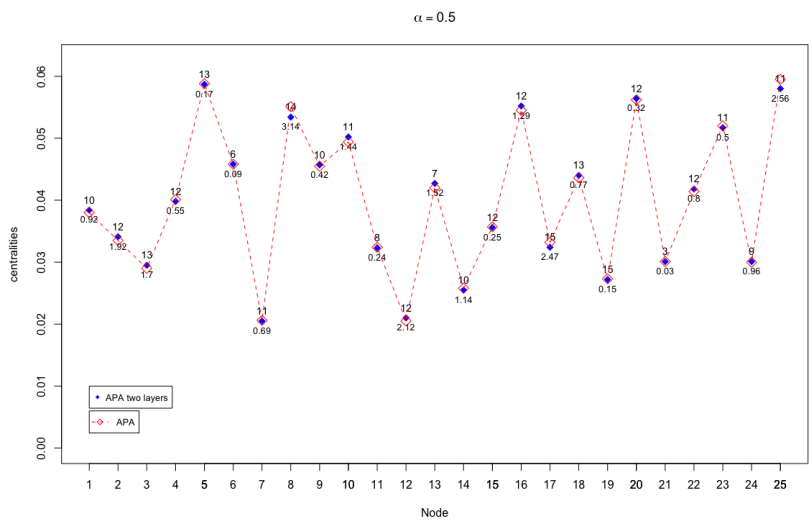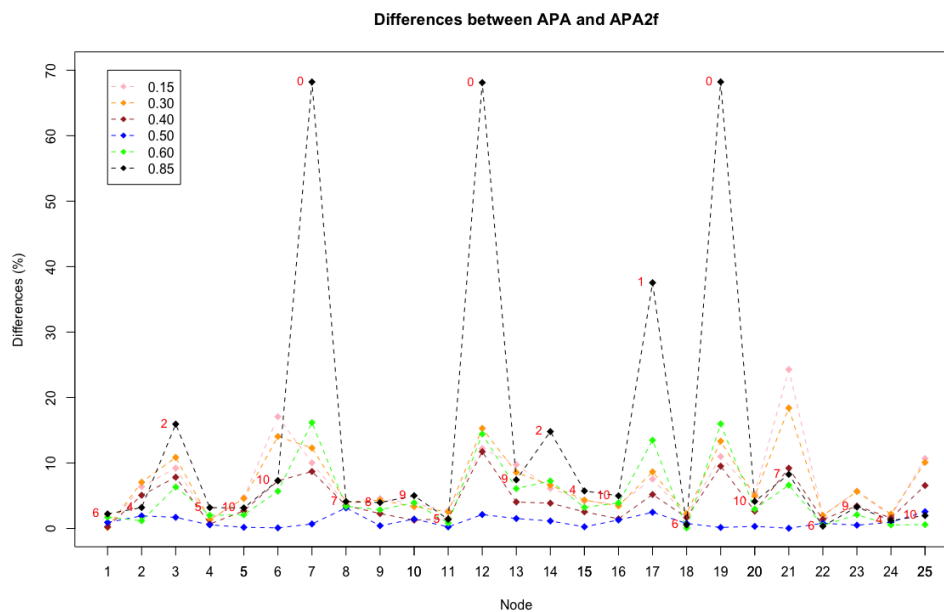

Figure 4. APA and APA2f centrality for $\alpha = 0.5$.

The case where there are more notable differences (up to 68%) occurs when $\alpha = 0.85$, that is, when the greatest importance to the data present in the network is given.

Now, let us study where and when these remarkable differences occur and the reasons for this behaviour. With this purpose, a new figure is constructed where the differences between the APA and APA2f centralities are reflected, for all the nodes in the network and for different values of the parameter.

Figure 5 shows the 25 nodes of the graph on the *X* axis and the differences, in percentages, between the APA and APA2f centrality measures on the *Y* axis. Lines of different colours are displayed for the chosen values of α that have been studied.

Figure 5. Differences between APA and APA2f centralities for different values of the parameter α.

The graphic shows the homogeneity of the line that represents the value $\alpha = 0.5$, which indicates the minimum differences existing in this case, in which exactly the same importance to the data and the topology of the network is assigned. Remark that the line is almost horizontal. Analysing the cases in which the parameter is kept close to the value 0.5, may be observed that the differences increase in some nodes, although they remain at maximum values that do not exceed 20%. Especially significant may be discussing the behaviour and patterns when $\alpha = 0.85$, which is the case where the greatest differences occur.

In order to extend the results to other types of networks, a real urban network is studied now to check if the results obtained for this particular type of spatial network follow a similar behaviour as those observed for randomly generated networks. Note that urban networks are characterized for a very uniform node degree distribution, since most of the nodes have degree in a range from 2 to 5. More exactly, the case study is an urban network representing a piece of the historic centre of the city of Murcia, Spain. The primal graph of the urban network is represented in Figure 6a. The graph has 185 nodes and 266 edges. The data vector associated with the nodes has been generated randomly, in order to carry out the appropriate comparisons.

In Figure 6b, the differences in the values of both centralities are summarised, for $\alpha = 0.85$.

Figure 7 shows the values of the APA and APA2f centralities for the 185 nodes of the urban network object of this study, taking $\alpha = 0.85$. The symbols in red colour represent the values of APA, while the symbols in blue colour represent the values of APA2f. In the upper part of the APA centrality the value of the data associated with each node is displayed, which is in a range from 0 to 10.

(**a**) Primal graph
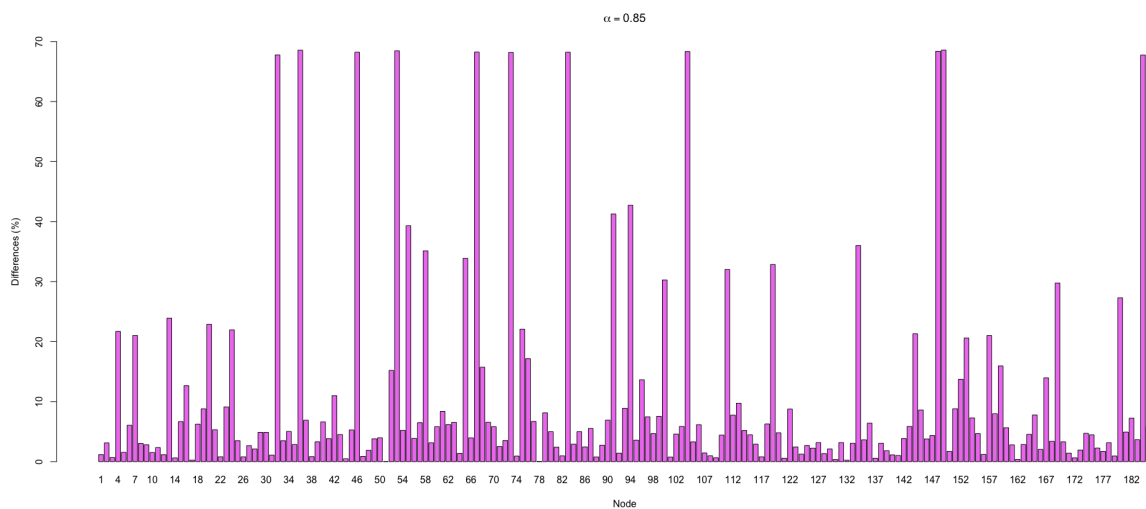
(**b**) Differences between APA and APA2f for $\alpha = 0.85$.

Figure 6. An example of a real urban network.

Likewise, in Figure 7 several features must be underlined. First, the smallest values of centrality correspond to the nodes that have a very low amount of data, which is reflected in the lower part of the figure. In contrast, in the upper part where the highest values of the centralities are produced, the nodes that have a greater amount of associated data can be found. This is in line with the idea of giving the greatest importance to the data taking, for instance, $\alpha = 0.85$.

Another remarkable pattern in this study is that the major differences in the values of the centralities occur in the extreme values, both the maximum and the minimum. For instance, for the nodes that have 0 or 1 data associated, which are those with a lower centrality, it is observed that the greatest differences occur in the APA and APA2f centralities with a well-defined pattern. The values of the APA2f centrality are lower than those of the APA centrality; the same pattern is repeated, that is, the APA2f centrality extend the range of centrality values. However, for the nodes with greater centralities corresponding to the maximum values of the data, APA2f is higher than APA.

A classic example in network science is the Zachary's karate club network. It constitutes a well-known social network of a university karate club described in [37]. The network captures

34 members of a karate club, documenting pairwise links between members who interacted outside the club (see Figure 8a). During the study of this network by Zachary, a conflict arose between the administrator and instructor, which led to the split of the club into two. Zachary correctly predicted each member's decision to belong to a certain group except for one member.
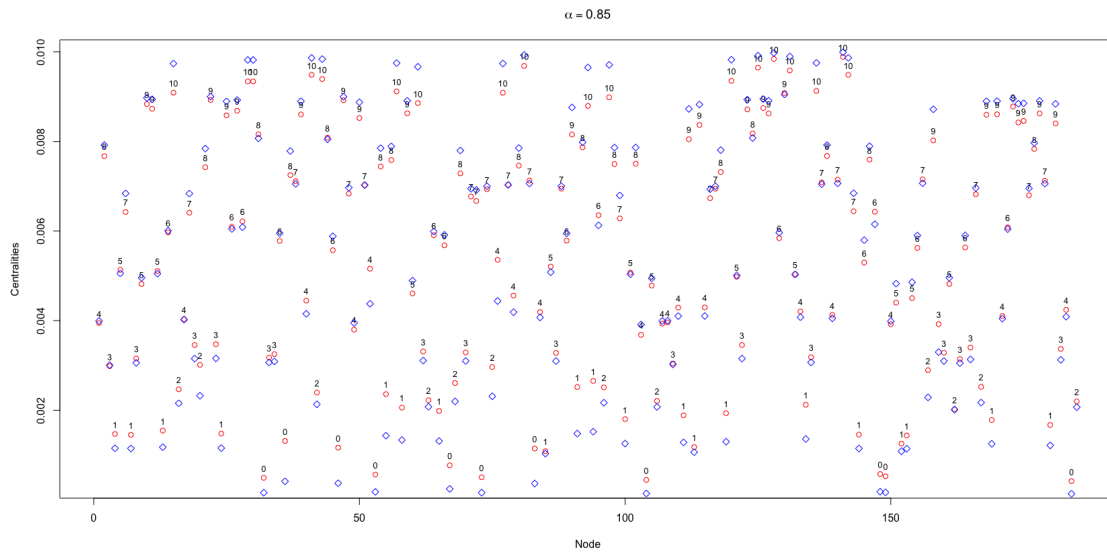

Figure 7. Centralities for the urban network in Figure 6a.

The relations between all the members of the club as a social network have been represented in Figure 8a. As it can be observed, the instructor (node 1) and president (node 34) are those that present a higher degree, that is, they are more connected with the rest of the club members.
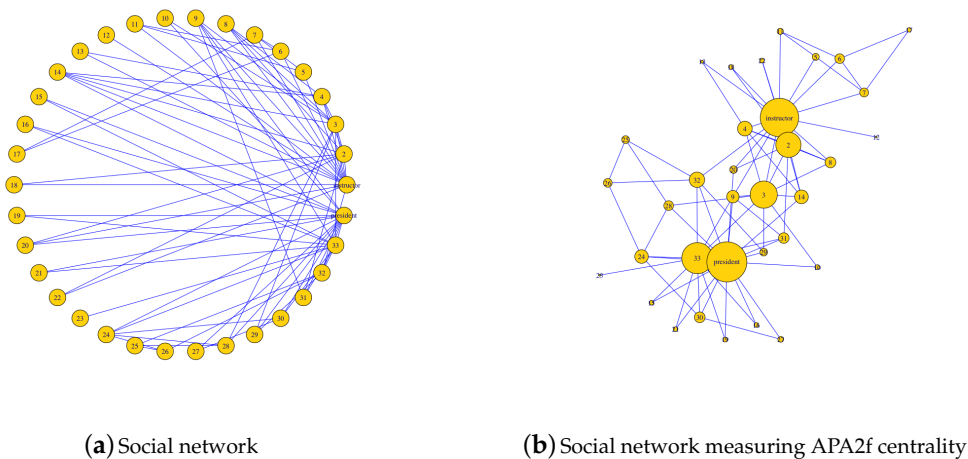

| (**a**) Social network | (**b**) Social network measuring APA2f centrality |

Figure 8. Applying APA2f centrality to Zachary's karate network.

To calculate the APA2f centrality a set of data associated with the nodes is required. The data used are fictitious and represent the number of messages that each club member receives on a specific day from the rest of the club members. These data have been generated proportional to the degree of each node, that is, they are in line with the degree of connectivity of each node. Specifically, the data vector used is:

$$D = [10, 8, 8, 4, 2, 3, 2, 3, 3, 1, 1, 0, 1, 4, 1, 1, 1, 1, 1, 2, 1, 1, 0, 4, 2, 3, 2, 2, 2, 3, 3, 4, 9, 10]^T .$$

The APA2f centrality has been calculated for this network and the result is the vector:

$$
\begin{aligned}
APA2f \quad = \quad & [0.0993\ 0.0652\ 0.0701\ 0.0379\ 0.0182\ 0.0250\ 0.0229\ 0.0281\ 0.0320\ 0.0125\ 0.0161\ 0.0046 \\
& 0.0071\ 0.0354\ 0.0122\ 0.0122\ 0.0114\ 0.0123\ 0.0122\ 0.0197\ 0.0122\ 0.0123\ 0.0049\ 0.0350 \\
& 0.0201\ 0.0226\ 0.0150\ 0.0251\ 0.0199\ 0.0277\ 0.0273\ 0.0397\ 0.0803\ 0.1035]
\end{aligned}
$$

Notice how the highest values of centrality correspond to the nodes 1 (instructor) and 34 (president), which are the most important people in the club. The centralities of each node have been displayed in Figure 8b. In this graph, the sizes of the nodes are proportional to the values of their centrality, resulting in the nodes with greater size the instructor and the president. Notice how two differentiated groups are observed in the club, according to the connections that each node has with the two main nodes.

4. Discussion

In this section, a general discussion of the results shown in the previous section by applying the APA2f centrality to different networks is carried out, paying special attention to the main findings and their implications as well as indicating the limitations of the proposal.

4.1. Comparing APA and APA2f

An essential point that must be addressed in the comparison between both measures is that there exists a coherence in the results that they offer when are applied to the same network. When talking about consistency in the results, it should not be understood that they must be exactly the same. The central issue is that APA and APA2f measures offer us a ranking of the network nodes, that is, a classification of the nodes in the network according to its importance within it. Therefore, the comparison should focus on two fundamental issues. On the one hand, the positions of the nodes in both rankings should not differ excessively. On the other hand, the results should not differ markedly because APA2f is based on APA. Therefore, may be interesting to carry out a numerical study about the differences that occur in both centralities when they are applied to graphs of different sizes.

The first objective raised with the comparison of both centralities is to check the coherence in the results obtained by measuring both centralities in any network. The coherence can be demonstrated by checking that both measures present a high correlation, in the sense that high values of one centrality correspond to high values of the other. Obviously, it would be a problem if a certain node were very important according to the APA centrality and irrelevant when the APA2f centrality is considered.

The discussion in the comparison should focus on the exhaustive analysis of the first numerical example of 25 nodes. Table 1 shows a coherence in the results between the two centralities, from the point of view of the classification of the nodes. The values computed by the two centralities are similar, except in some specific cases that will be analysed later.

Except for minor non-significant variations, the most important nodes are repeated in both centralities. No node with a high APA centrality has a low APA2f centrality and vice versa, the nodes with very low centralities according to the APA2f classification still have a very low centrality applying the APA model. For the validation of this new centrality APA2f, the behaviour of the correlation is very important.

To check the consistency of the results, the Spearman's rank correlation (denoted by ρ) is chosen, taking as variables the measures of the centralities APA and APA2f, respectively. Spearman's correlation is a non-parametric measure of rank correlation, that is, a coefficient that measure the statistical dependence between the rankings of two variables. A perfect Spearman' correlation will produce a coefficient $+1$. Intuitively, the Spearman's correlation between two variables will be high when observations have a similar or identical rank between the two variables, and low when observations have a dissimilar rank between the two variables.

Some tests calculating the Spearman's correlation have been developed, taking different values of the α parameter and sizes of the network (see Table 2).
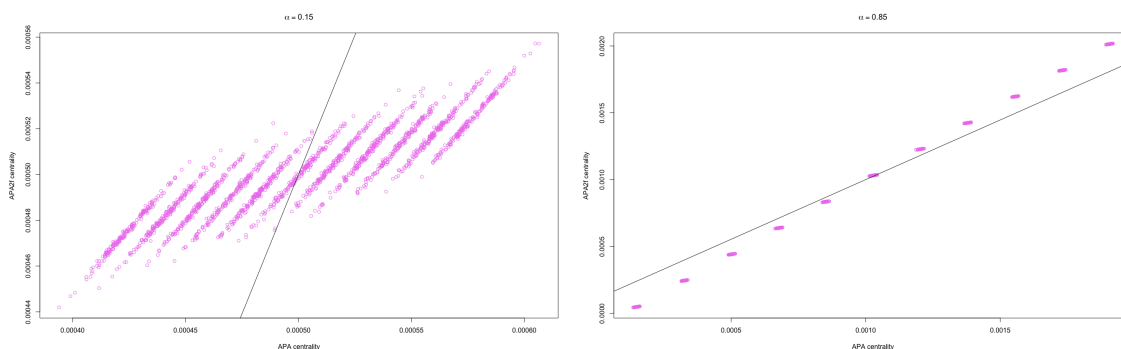
Table 2. Spearman's correlation for different values of the α parameter and different sizes of the network.

n	$\alpha = 0.15$	$\alpha = 0.30$	$\alpha = 0.50$	$\alpha = 0.85$
100	0.8131	0.9603	0.9995	0.9999
500	0.8413	0.9921	0.9999	0.9999
1000	0.8837	0.9959	0.9999	0.9999
2000	0.9181	0.9983	0.9999	0.9999

The results can be summarized in these key points:

- The values of the lowest Spearman's correlation between both variables occurs when $\alpha = 0.15$, regardless of the size of the matrix; these values are between 80% and 90%.
- The correlation when $\alpha = 0.5$ is 0.9999, which indicates a total correlation between the two centralities.
- When $\alpha > 0.5$, the correlation presents values higher than 99% for any size of the network.

In Figure 9, the APA versus APA2f Spearman's correlation for networks randomly generated with 2000 nodes have been displayed, for different values of the parameter: 0.15 (on the left) and 0.85 (on the right).

Figure 9. APA and APA2f Spearman's correlation for a 2000 nodes network.

The second objective in the comparison of both centralities was to analyse the magnitude of the differences in their values. To perform this task, let us carefully analyse Figures 3 and 5, as well as Table 1. There exist differences in the measures only for exceptional cases, especially when $\alpha = 0.85$. However, these differences do not modify substantially the order or classification of the nodes in terms of their importance within the network. This pattern is carefully analysed.

It may be remarked that for $\alpha = 0.5$ the coincidence in the results offered by both centralities is almost absolute. Analogously, as shown in Table 1, the first three most important nodes are the same. The differences found in the values of the centrality of these first three nodes are very small, which is not significant in their classification in order of importance.

When the parameter α moves away from the value 0.5, let us notice that the differences in the values of both centralities increase progressively, although the differences are only significant for some nodes. Analysing the cases in which the parameter is kept close to the value 0.5, it is observed that the differences increase in some nodes, although they remain at maximum values that do not exceed 20%. In Table 1, the greatest differences in percentage for each value of the parameter have been highlighted in red colour.

Especially significant may be analysing the case when $\alpha = 0.85$, where the greatest differences occur. The cases with a highest difference (up to 68%) occurs when $\alpha = 0.85$ in nodes $7, 12, 19$.

According to the APA design, $\alpha = 0.85$ means that the greatest importance to the data present in the network is given. The fact that these maximum differences occur in these nodes has an explanation by observing a pattern common to all of them. This pattern refers to the fact that these nodes do not have assigned data, so they can be considered as particular or singular nodes (especially when the data is given the greatest importance). As can be seen in Table 1, the next largest difference occurs in node 17, where the difference already drops to 40%. Note that node 17, the only that has associated 1 data (a very low value) is precisely the next one that presents notable differences.

According to this pattern, the biggest differences in the centralities occur with nodes that do not have data or have very few associated data and, in addition, the alpha parameter is close to 1.

It is important to remember that, although there are significant differences in the values of the centralities, for very specific cases with well-defined patterns, this does not imply an alteration in the classification of the nodes themselves. Analysing the numerical results it can be observed that the APA2f centrality extend the range of values of the measure.

Numerical results carried out with larger networks show us the same pattern that we have described for this example (see Table 2). The conclusion is that the results do not depend on the size of the matrix, as they are systematically repeated.

Regarding the example of the real urban network, it must be said that the behaviour of APA and APA2f centralities with respect to this real urban network does not differ with the networks studied previously. There exists a total coherence in the values of the centralities and the differences follow a similar pattern to that described for random networks. Again, the minimum differences occur when $\alpha = 0.5$, that is, when the topology and data are as important in the network. Similarly, the maximum differences are given when $\alpha = 0.85$ (greatest importance to data).

The graph shows how there is a group of nodes whose differences shoot around a concrete and constant value, such as 68%. Analysing the characteristics of this group of nodes, the same pattern as in the example studied is found: they all have 0 data associated. This correspondence is absolute, since there is no other node in the urban network with 0 data associated and that does not experience this maximum value of the difference of centralities. Let us emphasize again that the highest differences in the centrality values do not represent significant changes in the order or classification of the network nodes. Those nodes with maximum differences are also, as expected, the nodes with lower centralities.

Following with the same previous patterns, it is noteworthy that the following values in the differences fall to approximately 40%. Thus, for example, the nodes that have a single data associated, which represents a very low value, are found in the graph in the group that has differences between 30 and 40 percent.

The example of the karate club developed at the end of Section 3 presents a numerical behaviour similar to that analysed in other examples. The APA2f centrality allows us to establish a classification of the nodes according to their importance within the network. Moreover, the values of the new centrality are coherent with those of the APA centrality, establishing a Spearman's correlation coefficient of 0.9974 between both measures.

4.2. The APA2f Centrality

The comparison between the APA and APA2f centralities has allowed us to establish some properties and essential characteristics of the proposed new centrality.

Taking the main characteristics of APA centrality and the idea of understanding PageRank centrality by a two-layers approach, a new centrality is designed that presents the following positive aspects:

- It is possible to determine the centrality of a network considering a set of data present in it as determinants when assessing the importance of network nodes.
- The model is based, from the mathematical point of view, on the calculation of the dominant eigenvector of an irreducible and stochastic matrix by columns, for which its numerical stability is

assured. In addition, the spectral properties of the M_{APA2} matrix allow us to use the well-known *Power Method* for the calculation of the dominant eigenvector.

- The numerical results show that the range of values of APA2f centrality increases with respect to the values provided by the APA centrality. Especially, for spatial networks such as the example of the urban network; this represents an advantage in the aspect of the visualization of the network, since it will allow us a graphical representation with greater contrast when carrying out an interpolation of the values and their corresponding colour in the chosen gradient.

The main limitation we can find when working with this centrality is related to the characteristics and size of the M_{APA2} matrix. The APA centrality is based on the construction of a matrix of size $n \times n$ from which we obtain the dominant eigenvector that provides us the classification of the network nodes in order of importance. However, the two-layers approach to the PageRank model involves the construction of a block matrix whose size increases to $2n \times 2n$. This means that it is necessary to calculate the dominant eigenvector of a matrix double in size as in the case of the original PageRank. Consequently, for very large networks (hundreds of thousands of nodes) the computational requirements can be very high. In this case, a parallelization of the algorithm could be interesting, following the example of Migallón et al. [26]. However, the scope of application in which our research is developed is that of spatial networks, especially urban networks. This means that the size of networks when we represent cities by primal graphs is not too large (thousands of nodes).

Given that the new centrality is based on APA centrality and inherits from it certain positive characteristics related to numerical stability and the possibility of introducing data in the network, there are two aspects that must be taken into account for assess the use of the APA2f measure against APA. One of these aspects is related to the problem of data visualization. Our scope of application of complex networks focuses on spatial networks, mainly urban networks, where both the data and the positions of the nodes are geolocated. These networks present a limitation when it comes to visually representing centrality. The low degree distribution of the nodes means that the range of values of centrality is low, which hinders their visualization by means of basic interpolation techniques. However, as already mentioned, the APA2f centrality extends that range of values, especially in spatial networks with data. This should improve the final visualization of the network. Even more, it would be interesting to use, only for the visualization of the centrality, the eigenvector obtained directly from the M_{APA2} matrix, without normalizing, which will suppose to extend the range of centrality values notably.

Secondly, the centrality APA2f offers us a construction technique that may be used to design and implement new measures of centrality in *multilayer networks*. This centrality for multilayer networks would continue the initial idea of the APA model to consider networks with data, although the construction technique would be based on the APA2f model. All this is part of our future research lines.

When we consider the scope and possible beneficiaries of a model like the one proposed, the field is very broad and diverse. As an example, let us think about the example of Zachary's karate club. This network shows the relationships between the people who make up a karate club. Let us analyse the application of the APA2f centrality to this case. The model allows us to establish the importance of the nodes not only in terms of their connections or links but in terms of a set of data that can be diverse and from different sources. In our example, a dataset consisting on the set of messages received by each member of the club in one day has been considered. But it is possible to study their relationships under other perspectives by taking another dataset. Let us use a dataset associated with these members referring to their social or economic position, for instance, and consider the relationships between them in relation to this dataset. Or let us take data from different social networks and study their relationships.

This adaptability to data and the possibility of working with real or virtual data is the great advantage of a model like this. Let us also think about the possibilities that it offers us in terms of simulating the behaviour of the whole network when we make small or large variations in the data. Following with the karate club network, It is arguable that the members in the club in the best

economic position relate more to each other? As another example, suppose an urban network and consider the commercial activity developed in the city. Take food stores and supermarkets as dataset. It is possible to study the commercial activity of the city from these data. However, it is also possible to perform simulations by modifying the dataset. For instance, let us introduce a new supermarket in a geolocated position in the city and study how the commercial network is affected by this new data.

5. Conclusions

Everyday life of most people is characterised by the use of common or compatible technology for a wide range of personal, social, educational and business activities, and by the ability to transmit, receive and exchange digital data rapidly between places irrespective of distance. Therefore, if something defines our current society, it is the amount of data that is produced. This means that most of the complex networks that are studied or analysed are linked in an inseparable way to some type of data. This motivates the need to have measures of centrality to evaluate the importance of the nodes of a network so that they consider the inclusion of the data in it and its importance in the calculation of it.

Based on the APA centrality capable of measuring the importance of the nodes in a network with data following the idea of importance that applies the well-known PageRank centrality, we design and implement a new centrality measure APA2f that incorporates the idea of understanding the PageRank model as two-layers, that is, the physical or topological layer and the teleportation layer. In our case, we replace the teleportation layer with the layer where the data associated with the network are distributed.

The numerical results and experiments carried out with networks of different types and sizes make it possible to verify the coherence of APA2f with respect to the APA original, in the sense that the classifications of nodes in order of importance offered by both measures are similar. Given that the model has its origin in the PageRank vector concept, the alpha parameter that is inherited from this model now allows us to control the importance that is given in the calculation of the centrality to the two layers separately, that is, we can determine the importance of network connections and the influence of data itself simply by varying this parameter. A detailed study is performed for different α.

The numerical results also show that the APA2f centrality extends the range of values of APA for all the cases studied. Due to this numeric characteristic, from the point of view of visualization, APA2f allows us to carry out a clearer and more precise visualization of the networks following any of the usual techniques of interpolation. Another positive aspect of APA2f is that it also allows us to understand any network with data as a network composed of two separated layers, so there is a possibility of extension to multilayer networks.

Author Contributions: In the paper, all the authors have participated in a similar way.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Estrada, E. *The Structura of Complex Networks. Theory and Applications*; Oxford University Press: Oxford, UK, 2012.
2. Crucitti, P.; Latora, V.; Porta, S. Centrality measures in spatial networks of urban streets. *Phys. Rev. E* **2006**, *73*, 036125.
3. Bonacich, P. Simultaneous group and individual centrality. *Soc. Netw.* **1991**, *13*, 155–168.
4. Stephenson, K.; Zelen, M. Rethinking centrality: Methods and examples. *Soc. Netw.* **1989**, *11*, 1–37.
5. Wasserman, S.; Faust, K. *Social Network Analysis*; Cambridge University Press: Cambridge, UK, 1994.

6. Szell, M.; Lambiotte, R.; Thurner, S. Multirelational organization of large-scale social networks in an online world. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 13636–13641.

7. Freeman, L. Power-law distributions in empirical data. *SIAM Rev.* **1979**, *1*, 215–239.

8. Page, L.; Brin, S.; Motwani, R.; Winogrand, T. *The Pagerank Citation Ranking: Bringing Order to the Web*; Technical Report 1999-66; Stanford InfoLab: Stanford, CA, USA, 1999.

9. Calabrese, F.; Ratti, C.; Colonna, M.; Lovisolo, P.; Parata, D. Real-time urban monitoring using cell phones: A case study in Rome. *IEEE Trans. Intell. Transp. Syst.* **2011**, *25*, 141–151.

10. Chen, D.; Lu, L.; Shang, M.S.; Zhang, Y.; Zhou, T. Identifying influential nodes in complex networks. *Phys. A Stat. Mech. Appl.* **2012**, *391*, 1777–1787.

11. Zhou, Y.; Lu, L.; Li, M. Quantifying the influence of scientists and their publications: Distinguishing between prestige and popularity. *New J. Phys.* **2012**, *14*, 033033.

12. Crucitti, P.; Latora, V.; Porta, S. The network analysis of urban streets: A primal approach. *Plan. Des.* **2006**, *33*, 705–725.

13. Crucitti, P.; Latora, V.; Porta, S. The network analysis of urban streets: A dual approach. *Phys. A Stat. Mech. Appl.* **2006**, *369*, 853–866.

14. Jiang, B. Ranking spaces for predicting human movement in an urban environment. *Int. J. Geogr. Inf. Sci.* **2009**, *23*, 823–837.

15. Bloch, F.; Jackson, M.; Tebaldi, P. Centrality Measures in Networks. *arXiv* **2017**, arXiv:1608.05845.

16. Bonacich, P. Power and centrality: A family of measures. *Am. J. Sociol.* **1987**, *92*, 1170–1182.

17. Boldi, P.; Vigna, S. Axioms for centrality. *Internet Math.* **2014**, *10*, 222–262.

18. Freeman, L. A set of measures of centrality based on betweenness. *Sociometry* **1977**, *40*, 35–41.

19. Brandes, U. A faster algorithm for betweenness centrality. *J. Math. Sociol.* **2001**, *25*, 163–177.

20. Haveliwala, T. Topic-sensitive PageRank: A context-sensitive ranking algorithm for Web search. *IEEE Trans. Knowl. Data Eng.* **2003**, *15*, 784–796.

21. Sydow, M. Random surfer with back step. In Proceedings of the 13th International World Wide Web Conference, New York, NY, USA, 17–22 May 2004; pp. 352–253.

22. Berkhin, P. A survey on PageRank computing. *Internet Math.* **2005**, *2*, 73–120.

23. Garcia, E.; Pedroche, F.; Romance, M. On the localization of the personalized PageRank of complex networks. *Linear Algebra Appl.* **2013**, *439*, 640–652.

24. Langville, A.; Mayer, C. Deeper inside PageRank. *Internet Math.* **2005**, *1*, 335–380.

25. Bianchini, M.; Gori, M.; Scarselli, F. Inside PageRank. *ACM Trans. Internet Technol.* **2005**, *5*, 92–128.

26. Migallón, H.; Migallón, V.; Palomino, J.; Penadés, J. A heuristic relaxed extrapolated algorithm for accelerating PageRank. *Adv. Eng. Softw.* **2018**, *120*, 88–95.

27. Agryzkov, T.; Oliver, J.; Tortosa, L.; Vicent, J. An algorithm for ranking the nodes of an urban network based on the concept of PageRank vector. *Appl. Math. Comput.* **2012**, *219*, 2186–2193.

28. Agryzkov, T.; Tortosa, L.; Vicent, J. New Highlights and a New Centrality Measure Based on the Adapted PageRank Algorithm for Urban Networks. *Appl. Math. Comput.* **2016**, *291*, 14–29.

29. Agryzkov, T.; Tortosa, L.; Vicent, J.; Wilson, R. A centrality measure for urban networks based on the eigenvector centrality concept. *Environ. Plan. B* **2017**, 14–29, doi:10.1177/2399808317724444.

30. Conti, M.; Mohan, K. Opportunities in opportunistic computing. *Computer* **2010**, *43*, 42–55.

31. Zhang, Y.; Song, L.; Jiang, C.; Tran, N.; Dawy, Z.; Han, Z. A social-aware framework for efficient information dissemination in wireless ad hoc networks. *IEEE Commun. Mag.* **2017**, *55*, 174–179.

32. Boccaletti, S.; Bianconi, G.; Criado, R.; del Genio, C.I.; Gómez-Gardenes, J.; Romance, M.; Sendina-Nadal, I.; Wang, Z.; Zanin, M. The structure and dynamics of multilayer networks. *Phys. Rep.* **2014**, *544*, 1–122.

33. Boccaletti, S.; Bianconi, G.; Criado, R.; del Genio, C.I.; Gómez-Gardenes, J.; Romance, M.; Sendina-Nadal, I.; Wang, Z.; Zanin, M. Multilayer networks. *J. Complex Netw.* **2014**, *2*, 203–271.

34. Pedroche, F.; Romance, M.; Criado, R. A biplex approach to PageRank centrality: From classic to multiplex networks. *Chaos* **2016**, *26*, 065301.

35. Pedroche, F. Métodos de cálculo del vector PageRank. *Bol. Soc. Esp. Mat. Apl* **2007**, *39*, 7–30.

36. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2013.

37. Zachary, W. An Information Flow Model for Conflict and Fission in Small Groups. *J. Anthropol. Res.* **1977**, *33*, 452–473.