Similarity matrix average for aggregating multiplex networks

IOP Publishing J. Phys. Complex. 4 (2023) 025017 https://doi.org/10.1088/2632-072X/acda09

Journal of Physics: Complexity

OPEN ACCESS

RECEIVED
3 December 2022

REVISED
22 May 2023

ACCEPTED FOR PUBLICATION
30 May 2023

PUBLISHED
9 June 2023

PAPER

Similarity matrix average for aggregating multiplex networks

Federica Baccini[1,2,*], **Lucio Barabesi**[3] and **Eugenio Petrovich**[4,5]

[1] Department of Computer Science, University of Pisa, Pisa, Italy
[2] Institute for Informatics and Telematics, CNR, Pisa, Italy
[3] Department of Economics and Statistics, University of Siena, Siena, Italy
[4] Department of Philosophy and Education Sciences, University of Torino, Torino, Italy
[5] Department of Philosophy, Tilburg University, Tilburg, The Netherlands
[*] Author to whom any correspondence should be addressed.

E-mail: federica.baccini@phd.unipi.it

Keywords: multiplex network, similarity matrix, Jaccard coefficient, cosine similarity, SimRank, Fréchet mean, statistical journal network

Abstract

We introduce a methodology based on averaging similarity matrices with the aim of integrating the layers of a multiplex network into a single monoplex network. Multiplex networks are adopted for modelling a wide variety of real-world frameworks, such as multi-type relations in social, economic and biological structures. More specifically, multiplex networks are used when relations of different nature (layers) arise between a set of elements from a given population (nodes). A possible approach for analyzing multiplex similarity networks consists in aggregating the different layers in a single network (monoplex) which is a valid representation—in some sense—of all the layers. In order to obtain such an aggregated network, we propose a theoretical approach—along with its practical implementation—which stems on the concept of similarity matrix average. This methodology is finally applied to a multiplex similarity network of statistical journals, where the three considered layers express the similarity of the journals based on co-citations, common authors and common editors, respectively.

1. Introduction

Multilayer networks constitute an increasing active research topic with applications in many different disciplines, such as social or biological sciences. A multilayer network is a collection of individual networks—referred to as layers—each containing its own nodes and edges, with additional edges between the various layers. For more details on multilayer networks, see the recent monographs by Bianconi (2018), De Domenico (2022) and Dickison *et al* (2016), or the survey paper by Kivelä *et al* (2014).

A noteworthy special case of multilayer network is the so-called multiplex network, which is characterized by the same set of nodes in each layer (see e.g. Newman 2018, section 6.7). Indeed, in multiplex networks a unique node type is present, while different edge types may occur. In such a case, the linking edges between the various layers trivially connect replicas of the same node in the different layers, even if such edges (interlayers) are omitted in practice for simplicity. Characteristic examples of multiplex networks are social networks, where the nodes are the individuals in a well-defined community with different types of relational ties between them (such as friendship, family or co-worker connections). Each type of relation is represented as a separate layer—see Bianconi (2018) for real examples of social, on-line social, economic and financial multiplex networks. A multiplex structure often emerges when studying relations among a fixed population of n objects which are described by m different sets of features. This setting can be modelled as an m-layered network, where each layer is a bipartite (two-mode) network and one set of the bipartition (the set of n objects) is common to all the layers, while the other varies between layers (each of the m sets of features). Thus, one can consider the one-mode projections on the common set of nodes on each layer, or a suitable similarity score between the objects for each type of feature. This framework produces a multiplex

(similarity) network, where the nodes represent the set of n objects. A multiplex network illustration in the field of scientometrics is provided by Baccini *et al* (2022). Specifically, these authors consider multiplex networks of journals in different disciplines constituted by three layers based on co-citations, common authors, and common editors (see also Baccini *et al* 2020). A further instance of multiplex network modelling can be found in the field of epigenetics, as shown by Baccini *et al* (2022). Indeed, Baccini *et al* (2022) consider a multiplex network of blood cell phenotypes, where the layers are based on different epigenetic modifications. In both the examples, the multiplex is obtained after the computation of a similarity score starting from a multilayer network with bipartite networks on each layer, as previously described.

A tool for investigating the structure of multiplex networks is based on the 'aggregation' (in some appropriate way) of the different layers in order to obtain a suitable monoplex network, i.e. a single layer. The integration of the information contained in multilayer networks, including the special case of multiplex networks, is motivated by the need of facilitating the visualization and the analysis of the global structure of a set of interacting entities. Indeed, if the aggregated network is a suitable representation of the global structure, it can reveal properties that are hardly detectable from the multi-layered structure (Wang *et al* 2014). This research direction is particularly relevant in genetic and biology, where several methods have been developed to integrate multi-omics data (see, for instance, Argelaguet *et al* 2018, 2020, Huizing *et al* 2023). Moreover, the resulting monoplex network is useful to interpret the strong connections between nodes and—eventually—to implement a cluster analysis for searching cohesive groups in the original multiplex network. In addition, the analysis of the correlation between the monoplex and the single layers may be helpful to detect the global community structure.

In such a setting, the target is focused on how to aggregate the different layers. A straightforward procedure for implementing an aggregated network may be based on a weighted average of the adjacency matrices corresponding to the layers (see e.g. Kivelä *et al* 2014, and references therein). Even if this strategy actually produces a single weighted adjacency matrix which characterizes the monoplex, it is prone to some drawbacks, since some layers may be over-represented. Moreover, the choice of the weights is subjective (Kivelä *et al* 2014). A more proper approach would be to consider similarity matrices—with suitable features—obtained from the adjacency matrices corresponding to the single layers, and then carry out an 'average' of these similarity matrices according to well-defined mathematical properties. Subsequently, by means of an appropriate similarity matrix average, the aggregated network is obtained.

In the present paper, we consider some recent advances in the theory of barycenters of objects lying in abstract spaces finalized to the concept of average for positive definite matrices (Bhatia 2009, Álvarez-Esteban *et al* 2016). More precisely, starting from two-mode multilayer networks, we first show that the corresponding similarity matrices based on the commonly-adopted similarity measures—such as the Jaccard, the cosine, and the SimRank similarity—belong to the space of completely positive matrices, which is a subspace of positive definite matrices. The space of completely positive matrices is the very natural space for similarity matrices, since its elements are the symmetric positive definite matrices with nonnegative entries. Subsequently, in order to aggregate multiplex networks, we consider the Fréchet mean criterion (see e.g. Bacák 2014) with three different metric choices—i.e. the classical Frobenius metric, the Riemannian metric, and the Wasserstein metric—to achieve appropriate definitions of average for positive definite matrices. In addition, we remark the properties of the averages obtained with the three metrics. We also discuss the choice of the weights to be adopted in the Fréchet criterion. Finally, we illustrate the theoretical findings with an application to the aforementioned multiplex networks considered by Baccini *et al* (2022) in the field of scientometrics.

The paper is organized as follows. Section 2 contains some preliminaries and notation. In section 3 we show that similarity matrices based on Jaccard, cosine and SimRank similarity are completely positive. In section 4 we present the theoretical framework to compute the averages of the similarity matrices connected to the multiplex networks. In section 5 we propose some choices for the weights involved in the averages. Section 6 presents the details of the code implementation for the practical computation of the averages. Section 7 presents the application to the multiplex network involving statistical journals. Finally, section 8 draws some conclusions.

2. Some notations and preliminaries

Let us assume that $\mathcal{P}_n$ is the space of symmetric positive semidefinite matrices of order n, i.e.

$$\mathcal{P}_n = \{\boldsymbol{X} \in \mathbb{R}^{n\times n} : \boldsymbol{X} = \boldsymbol{X}^{\mathrm{T}}, \boldsymbol{X} \succeq \boldsymbol{0}\},$$

where $\boldsymbol{X} \succeq \boldsymbol{0}$ denotes that the eigenvalues of $\boldsymbol{X}$ are nonnegative. The space $\mathcal{P}_n$ is a differentiable manifold in the set of Hermitian matrices (for more details on this class of matrices, see Bhatia 2009, chapter 6). The

Frobenius inner product on $\mathcal{P}_n$ is defined by $\langle \boldsymbol{X},\boldsymbol{Y}\rangle_{\mathrm{F}} = \mathrm{tr}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y})$ for $\boldsymbol{X},\boldsymbol{Y} \in \mathcal{P}_n$, and the associated Frobenius norm is given by $\|\boldsymbol{X}\|_{\mathrm{F}} = \mathrm{tr}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{1/2}$. If $\boldsymbol{X} \in \mathcal{P}_n$, $\boldsymbol{X}^{1/2}$ denotes the unique positive definite matrix such that $\boldsymbol{X} = \boldsymbol{X}^{1/2}\boldsymbol{X}^{1/2}$, while the corresponding normalized matrix is given by $\lambda_1(\boldsymbol{X})^{-1}\boldsymbol{X}$, where $\lambda_1(\boldsymbol{X}) > 0$ represents the largest eigenvalue of $\boldsymbol{X}$. Finally, if $\boldsymbol{X} = (x_{ij})$ is a square matrix of order n, we assume that $\mathrm{diag}(\boldsymbol{X}) = (\delta_{ij}x_{ij})$, while if $\boldsymbol{x} = (x_1,\ldots,x_n)^{\mathrm{T}}$ is a vector then $\mathrm{diag}(\boldsymbol{x}) = (\delta_{ij}x_i)$, where δ_{ij} represents the Kronecker delta.

Let us consider the subspace $\mathcal{CP}_n \subset \mathcal{P}_n$ of completely positive matrices, i.e.

$$\mathcal{CP}_n = \{\boldsymbol{X} \in \mathbb{R}^{n\times n} : \boldsymbol{X} = \boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y}, \exists \boldsymbol{Y} \in \mathbb{R}^{p\times n}, \boldsymbol{Y} \geqslant \boldsymbol{0}\},$$

where $\boldsymbol{X} \geqslant \boldsymbol{0}$ denotes that the entries of $\boldsymbol{X}$ are nonnegative. For the properties of this class of matrices, see the monographs by Berman and Shaked-Monderer (2003), Shaked-Monderer and Berman (2021) and Johnson *et al* (2020). A necessary condition for a matrix $\boldsymbol{X}$ to be completely positive is that $\boldsymbol{X} \succeq \boldsymbol{0}$ and $\boldsymbol{X} \geqslant \boldsymbol{0}$, even if—contrary to intuition—this condition is not generally sufficient (Berman and Shaked-Monderer 2003). In addition, Berman and Shaked-Monderer (2003) show that $\mathcal{CP}_n$ is a closed convex cone in the class of Hermitian matrices—and hence in $\mathcal{P}_n$. The following proposition provides some results on completely positive matrices which will be helpful in the following sections.

Proposition 2.1. *By assuming that $\boldsymbol{X},\boldsymbol{Y} \in \mathcal{CP}_n$, it holds:*

(i) $\boldsymbol{X}+\boldsymbol{Y} \in \mathcal{CP}_n$;
(ii) $\boldsymbol{X}\odot\boldsymbol{Y} \in \mathcal{CP}_n$, *where the symbol $\odot$ denotes the Hadamard product;*
(iii) $\boldsymbol{Z}^T\boldsymbol{X}\boldsymbol{Z} \in \mathcal{CP}_n$ *if* $\mathbf{Z}$ *is a square matrix of order* n *such that* $\boldsymbol{Z} \geqslant \boldsymbol{0}$.

Proof. See corollary 2.1, 2.2 and proposition 2.2 by Berman and Shaked-Monderer (2003). □

3. Completely positive similarity matrices from two-mode networks

In the present section we discuss the emergence of similarity matrices from the one-mode projection of a two-mode network. In particular, we show that the resulting matrices are completely positive. Let us consider two-mode (bipartite) networks, i.e. networks displaying two types of nodes where edges solely tie nodes of different type (for more details on bipartite networks, see e.g. Newman 2018, chapter 6). This kind of networks is commonly adopted to describe the membership of a set of p items to n groups. The items are represented by the first set of nodes, while the groups are represented by the second set of nodes—and the edges connect the items to the groups they belong to. A bipartite network captures exactly the same information as a hypergraph, even if for most purposes bipartite networks are more convenient (Battiston *et al* 2020). Indeed, many social and biological networks belong to this family. For instance, in social sciences, bipartite graphs are suitable to model the membership of individuals to groups of variable size, or to describe a collaboration of authors to scientific papers. In biology, a protein-protein interaction network can be seen as a two-mode network where proteins are linked to the protein complexes they belong to (for more examples of real-world bipartite networks see e.g. Wasserman and Faust 1994 and Pavlopoulos *et al* 2018).

A two-mode network can be characterized by an incidence matrix $\boldsymbol{B} = (b_{ij})$ of order $(p \times n)$, where

$$b_{ij} = \begin{cases} 1 & \text{if item } i \text{ belongs to group } j \\ 0 & \text{otherwise .} \end{cases}$$

To avoid triviality, we assume that there exist $i \in \{1,\ldots,p\}$ such that $b_{ij} = 1$ for each $j = 1,\ldots,n$, i.e. there is at least one item for each group. Obviously, it holds that $\boldsymbol{B} \geqslant \boldsymbol{0}$. In order to provide a practical illustration in the scientometric setting, the network of journals and editors in a given discipline may be considered as an example of a two-mode bipartite network with p editors as items and n journals as groups. In such a case, b_{ij} constitutes the indicator variable for the membership of the ith editor to the jth journal. On the basis of the incidence matrix $\boldsymbol{B}$, the one-mode projection with respect to groups of the two-mode bipartite network can be considered. Hence, if $\boldsymbol{G} = (g_{ij})$ is the matrix of order $(n \times n)$ such that $\boldsymbol{G} = \boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}$, it holds that $g_{ij} = \sum_{k=1}^{p} b_{ki}b_{kj}$, i.e. the entry g_{ij} gives the total number of items which are in both the ith group and the jth group for $i,j = 1,\ldots,n$. Consequently, g_{ii} gives the number of items in the ith group. As an example, in the network of journals and editors, g_{ij} provides the number of editors belonging to the board of both the ith journal and the jth journal, while g_{ii} is the number of editors in the board of the ith journal. This definition of $\boldsymbol{G}$ implies that $\boldsymbol{G} \in \mathcal{CP}_n$.

The weighted adjacency matrix $\boldsymbol{Z} = (z_{ij})$ of order $(n \times n)$ corresponding to the weighted one-mode projection is defined as $\boldsymbol{Z} = \boldsymbol{G} - \text{diag}(\boldsymbol{G})$. Thus, z_{ij} denotes the weight of the edge between the ith group node and the jth group node. The corresponding unweighted adjacency matrix $\boldsymbol{A} = (a_{ij})$ of order $(n \times n)$ is such that $a_{ij} = \mathbb{1}_{\{z_{ij}>0\}}$ for $i,j = 1,\dots,n$, where $\mathbb{1}_E$ is the indicator function of the set E. Hence, a_{ij} is binary-valued, as it assumes value 1 if there is an edge between the ith group node and the jth group node, 0 otherwise. As an example, in the network of journals and editors, $a_{ij} = 1$ if the ith journal and the jth journal share at least one editor, and $a_{ij} = 0$, otherwise. It is clear that $\boldsymbol{Z} \geqslant \boldsymbol{0}$ and $\boldsymbol{A} \geqslant \boldsymbol{0}$. Finally, we consider the matrix $\boldsymbol{F} = (f_{ij})$ given by $\boldsymbol{F} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{A}^2$. Thus, since $f_{ij} = \sum_{k=1}^{n} a_{ki}a_{kj}$, the entry f_{ij} gives the total number of common neighbors between the ith group node and the jth group node for $i,j = 1,\dots,n$. Moreover, since $a_{ij}^2 = a_{ij}$ for $i,j = 1,\dots,n$, the entry f_{ii} gives the degree corresponding to the ith group node. On the basis of its definition, it is also follows that $\boldsymbol{F} \in \mathcal{CP}_n$.

The introduction of the matrices $\boldsymbol{G}$, $\boldsymbol{Z}$ and $\boldsymbol{F}$ permits to analyze some common choices of similarity measures that allow to build one-mode networks. Our aim is to show that the commonly adopted similarity matrices are completely positive. Newman (2018, section 7.6) indicates several measures of structural and regular equivalence, which are appropriate for assessing the similarity of network nodes. Concerning the normalized similarity measures of structural equivalence, g_{ij} belongs itself to this class of similarity measures. Therefore, $\boldsymbol{G}$ may be considered as a similarity matrix, even if it may be not suitable to adopt this option in practice. Some authors implement the similarity matrices of structural equivalence on the basis of the matrix $\boldsymbol{G}$ (see e.g. Leydesdorff 2008), while others suggest the matrix $\boldsymbol{F}$ to the same aim (see e.g. Newman 2018, section 7.6). Since both these matrices are completely positive, we provide the following results by adopting $\boldsymbol{G}$—even if they can be also obtained by using $\boldsymbol{F}$ in place of $\boldsymbol{G}$.

In the context of similarity matrices of structural equivalence, Newman (2018) emphasizes that the well-known Jaccard coefficient and the cosine similarity coefficient are the most widely used for the practical analysis of networks (see also Van Eck and Waltman 2009). We first consider the similarity matrix $\boldsymbol{J} = (J_{ij})$ of order $(n \times n)$ which is based on the Jaccard coefficient. More precisely, J_{ij} constitutes the Jaccard coefficient between the ith group and the jth group, i.e.

$$J_{ij} = \frac{g_{ij}}{g_{ii} + g_{jj} - g_{ij}}.$$

It should be remarked that the quantity $(g_{ii} + g_{jj} - g_{ij})$ in the denominator of J_{ij} actually represents the total number of items belonging to the ith group or to the jth group. Hence, J_{ij} is a normalized similarity measure in $[0,1]$ obtained by considering the common items to the two groups divided by the total number of distinct items in the two groups. Since $\boldsymbol{G}$ is symmetric, it can be easily proved that $\boldsymbol{J}$ is in turn symmetric. The following proposition provides the target result on $\boldsymbol{J}$.

Proposition 3.1. *The matrix* $\mathbf{J}$ *is completely positive.*

Proof. Let us consider the matrix $\boldsymbol{U} = (u_{ij})$ of order $(n \times n)$ where

$$u_{ij} = p - g_{ii} - g_{jj} + g_{ij},$$

i.e. u_{ij} gives the total number of items which do not belong to the ith group and/or to the jth group. Since $\boldsymbol{G} = \boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}$ and since $b_{ij}^2 = b_{ij}$ for $i,j = 1,\dots,n$, it reads

$$u_{ij} = \sum_{k=1}^{p}(1 - b_{ki} - b_{kj} + b_{ki}b_{kj}) = \sum_{k=1}^{p}(1 - b_{ki})(1 - b_{kj}).$$

Hence, it holds $\boldsymbol{U} = (\mathbf{1}_p\mathbf{1}_n^{\mathrm{T}} - \boldsymbol{B})^{\mathrm{T}}(\mathbf{1}_p\mathbf{1}_n^{\mathrm{T}} - \boldsymbol{B})$, where $\mathbf{1}_p$ denotes the column vector of length p with all entries equal to 1, and $\mathbf{1}_n^{\mathrm{T}}$ is the row vector of length n with elements all equal to 1. Since $\mathbf{1}_p\mathbf{1}_n^{\mathrm{T}} - \boldsymbol{B} \geqslant \boldsymbol{0}$, we have that $\boldsymbol{U} \in \mathcal{CP}_n$. Moreover, let us consider the matrix $\boldsymbol{V} = (v_{ij})$ of order $(n \times n)$ such that

$$v_{ij} = \frac{1}{1 - u_{ij}/p} = \sum_{k=0}^{\infty}\frac{1}{p^k}u_{ij}^k,$$

where the last equality is obtained from the Geometric series, since it holds $0 < u_{ij} < p$ on the basis of the definition of u_{ij}. The matrix $\boldsymbol{V}$ may be expressed as

$$\boldsymbol{V} = \sum_{k=0}^{\infty}\frac{1}{p^k}\boldsymbol{U}_k,$$

where the matrix $\boldsymbol{U}_k$ is recursively defined as $\boldsymbol{U}_k = \boldsymbol{U}_{k-1} \odot \boldsymbol{U}$ with the initial condition $\boldsymbol{U}_0 = \mathbf{1}_n \mathbf{1}_n^{\mathrm{T}}$. Since the Hadamard product of completely positive matrices is in turn a completely positive matrix from (*ii*) of proposition 2.1 and $\boldsymbol{U}_0 \in \mathcal{CP}_n$, it also holds that $\boldsymbol{U}_k \in \mathcal{CP}_n$. In addition, since also the sum of completely positive matrices is a completely positive matrix from (*i*) of proposition 2.1, and $\mathcal{CP}_n$ is a closed cone, it follows that $\boldsymbol{V} \in \mathcal{CP}_n$. Thus, since

$$\boldsymbol{J} = \frac{1}{p}\boldsymbol{G} \odot \boldsymbol{V}$$

and $\boldsymbol{G} \in \mathcal{CP}_n$, we conclude that $\boldsymbol{J} \in \mathcal{CP}_n$. □

In the framework of structural equivalence, a further suitable similarity matrix is obtained from the cosine similarity coefficient. This choice yields a similarity matrix $\boldsymbol{C} = (c_{ij})$ of order $(n \times n)$, where the entry c_{ij} represents the cosine similarity coefficient

$$c_{ij} = \frac{g_{ij}}{(g_{ii}g_{jj})^{1/2}}. \tag{1}$$

The denomination 'cosine coefficient' derives from its morphology as a ratio of an inner product of two vectors to the product of the corresponding norms. Hence, on the basis of the Cauchy–Schwarz inequality, c_{ij} is a normalized similarity measure in $[0, 1]$. Moreover, since $\boldsymbol{G}$ is symmetric, it can be easily shown that $\boldsymbol{C}$ is in turn symmetric. The following proposition gives a result analogous to proposition 3.1 for the matrix $\boldsymbol{C}$.

Proposition 3.2. *The matrix* $\mathbf{C}$ *is completely positive.*

Proof. Let us consider the vector $\boldsymbol{v} = (g_{11}^{-1/2}, \ldots, g_{nn}^{-1/2})^{\mathrm{T}}$ of order n and the diagonal matrix $\boldsymbol{U} = \mathrm{diag}(\boldsymbol{v})$. The matrix $\boldsymbol{C}$ can be expressed as

$$\boldsymbol{C} = \boldsymbol{U}^{\mathrm{T}}\boldsymbol{G}\boldsymbol{U} = \boldsymbol{U}^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{U} = (\boldsymbol{B}\boldsymbol{U})^{\mathrm{T}}\boldsymbol{B}\boldsymbol{U},$$

and hence $\boldsymbol{C} \in \mathcal{CP}_n$ since $\boldsymbol{B}\boldsymbol{U} \geqslant \mathbf{0}$. □

In the framework of normalized similarity measures of regular equivalence, the SimRank similarity measure, originally introduced by Jeh and Widom (2002), is often considered (see Newman 2018, section 7.6). The SimRank similarity matrix $\boldsymbol{\Sigma} = (\sigma_{ij})$ of order $(n \times n)$ satisfies the matrix equation $\boldsymbol{\Sigma} = c\boldsymbol{P}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{P} + \boldsymbol{D}$, where $c \in (0, 1)$ is the so-called delay factor, while $\boldsymbol{D} = \mathrm{diag}(\boldsymbol{d})$ with $\boldsymbol{d} = (d_1, \ldots, d_n)^{\mathrm{T}}$ is such that $d_i \in [1 - c, 1]$ for $i = 1, \ldots, n$ (see e.g. Liao *et al* 2019). In addition, $\boldsymbol{P} = (p_{ij})$ is the column-normalized adjacency matrix $\boldsymbol{A}$, i.e. $p_{ij} = a_{ij} / \sum_{l=1}^{n} a_{il}$, and hence $\boldsymbol{P} \geqslant \mathbf{0}$. In turn, the following proposition provides the target result for $\boldsymbol{\Sigma}$.

Proposition 3.3. *The matrix* $\boldsymbol{\Sigma}$ *is completely positive.*

Proof. The matrix equation which defines $\boldsymbol{\Sigma}$ may be recursively solved in order to obtain the series representation

$$\boldsymbol{\Sigma} = \sum_{k=0}^{\infty} c^k (\boldsymbol{P}^{\mathrm{T}})^k \boldsymbol{D}\boldsymbol{P}^k.$$

Since $\boldsymbol{P} \geqslant \mathbf{0}$ by definition, it holds that $\boldsymbol{P}^k \geqslant \mathbf{0}$ on the basis of (*ii*) of proposition 2.1 and hence $(\boldsymbol{P}^{\mathrm{T}})^k \boldsymbol{D}\boldsymbol{P}^k = (\boldsymbol{P}^k)^{\mathrm{T}} \boldsymbol{D}\boldsymbol{P}^k \in \mathcal{CP}_n$ on the basis of (*iii*) of proposition 2.1. Since the sum of completely positive matrices is a completely positive matrix from (*i*) of proposition 2.1 and $\mathcal{CP}_n$ is a closed cone, it follows that $\boldsymbol{\Sigma} \in \mathcal{CP}_n$. □

In this section we have dealt with the Jaccard and cosine similarity, as well as the SimRank (strictly connected to the Katz similarity), since they are usually considered as the most representative and the most frequently-used similarity measures in the framework of structural and regular equivalence (see Newman 2018, and references therein). The preference between these similarities mostly depends on the nature of the data to be analyzed. In any case, despite the choice of the similarity measure, it should be generally required that the corresponding similarity matrix is completely positive, since $\mathcal{CP}_n$ is the natural space for such a matrix.

4. Aggregation of multiplex networks

Let us consider a multilayer network where each layer is a two-mode bipartite network. We assume the additional property that the bipartite graphs on each layer have in common one of the two sets of the node bipartition. The distinct subsets of nodes can be interpreted as distinct sets of items, while the vertex subset can be seen as a set of groups the different items belong (or do not belong) to. A practical example of this structure is constituted by a network of journals in a given discipline with respect to the editors in the journal boards and to the authors contributing to the journals. These relations can be modelled as a two-layer network of the type described above. Indeed, the first layer is a bipartite graph connecting editors (items) to journals (groups); the second layer is a bipartite graph connecting authors (items) to journals (groups). Therefore, the set of journals (groups) is common to both layers. This particular multilayer structure can be turned into a multiplex network by considering the one-mode projection of each layer on the common subset of nodes. Figure 1 displays the example of the simple two-layer journal network just discussed (figure 1(a)), which can be turned into a multiplex network of journals (figure 1(b)) through one-mode projection.

Let us deal with a multilayer network with m layers having the structure described above. For the seek of simplicity, we will refer to the common set of nodes in all the layers as the set of groups, and to the distinct set of nodes as the set of items. Let n denote the number of groups. Formally, m two-mode bipartite networks are considered, where for the kth layer there exists an incidence matrix $\boldsymbol{B}_k = (b_{k,ij})$ of order $(p_k \times n)$, with p_k denoting the set of items in the kth layer, and

$$b_{k,ij} = \begin{cases} 1 & \text{if item } i \text{ of layer } k \text{ belongs to group } j \\ 0 & \text{otherwise}, \end{cases}$$

with $k = 1, \ldots, m$. Thus, there are m matrices $\boldsymbol{G}_k = (g_{k,ij})$ of order $(n \times n)$ given by $\boldsymbol{G}_k = \boldsymbol{B}_k^{\mathrm{T}} \boldsymbol{B}_k$ for $k = 1, \ldots, m$. Correspondingly, there are k weighted and unweighted adjacency matrices $\boldsymbol{Z}_k = (z_{k,ij})$, $\boldsymbol{A}_k = (a_{k,ij})$ and $\boldsymbol{F}_k = (f_{k,ij})$ of order $(n \times n)$ for $k = 1, \ldots, m$. Based on the matrices $\boldsymbol{G}_k$s or $\boldsymbol{F}_k$s, it is possible to define m suitable similarity matrices for each layer. Thus, we explore the aggregation of m layers of the multiplex network into a monoplex, i.e. we assess how to achieve a suitable synthesis of the multiplex network. This problem is pursued by averaging in some appropriate way the m similarity matrices corresponding to the layers, which will be denoted as $\boldsymbol{S}_1, \ldots, \boldsymbol{S}_m$. Specifically, three proposals for a suitable average, which will be indicated as $\boldsymbol{S}_+$, of the m similarity matrices $\boldsymbol{S}_1, \ldots, \boldsymbol{S}_m \in \mathcal{CP}_n$ will be considered.

The computation of the barycenter of a set of objects is generally based on the minimization of an appropriate objective function. More precisely, $\boldsymbol{S}_+$ is obtained as

$$\boldsymbol{S}_+ = \arg\min_{\boldsymbol{X} \in \mathcal{P}_n} \sum_{k=1}^{m} w_k d(\boldsymbol{X}, \boldsymbol{S}_k)^2, \tag{2}$$

where $d : \mathcal{P}_n \times \mathcal{P}_n \to \mathbb{R}^+$ is a metric on $\mathcal{P}_n$, while the known weights $w_1, \ldots, w_m$ are such that $w_k \geqslant 0$ for $k = 1, \ldots, m$ and $\sum_{k=1}^{m} w_k = 1$. In a general framework, $\boldsymbol{S}_+$ provides the so-called Fréchet mean (see e.g. Bacák 2014). However, the Fréchet mean might be not suitable in an arbitrary metric space, although it is highly appropriate in a geodesic metric space of non-positive curvature—a Hadamard space—such as $\mathcal{P}_n$. The existence and uniqueness of the minimizer $\boldsymbol{S}_+$ in a Hadamard space is assured by theorem 2.4 by Bacák (2014). It should be remarked that $\boldsymbol{S}_+$ is defined as a minimizer on $\mathcal{P}_n$ in expression (2). However, we seek for a $\boldsymbol{S}_+$ that is, in turn, a similarity matrix, i.e. we would require that $\boldsymbol{S}_+$ is symmetric positive semidefinite with non-negative entries. Since $\mathcal{CP}_n$ is a closed convex cone in $\mathcal{P}$ (see section 2), the existence and uniqueness of $\boldsymbol{S}_+$ in $\mathcal{CP}_n$ are guaranteed if the m similarity matrices $\boldsymbol{S}_1, \ldots, \boldsymbol{S}_m$ are defined in $\mathcal{CP}_n$.

We have to point out that the proposed statistical approach, which is based on the minimization of expression (2), is descriptive in its own nature, since we do not assume a stochastic network setting. Indeed, our target consists in providing a reasonable concept of network barycenter when a multiplex network is considered for exploratory data analysis. This goal is fulfilled by means of the minimization of a weighted sum of squared matrix distances, in the spirit of the celebrated least squares method. However, advanced inferential frameworks exist, where statistical models on network spaces are suitably indexed by an average population network (to be appropriately defined). For example, by considering a network space equipped with the Hamming distance, Banks and Carley (1994) proposed an exponential-type distribution for network-valued random mappings characterized by a network-valued 'location' parameter (defined as the central network by these authors) and a real-valued 'dispersion' parameter. Noteworthy, the maximum likelihood estimate of the central network corresponds to the sample Fréchet mean, as given in

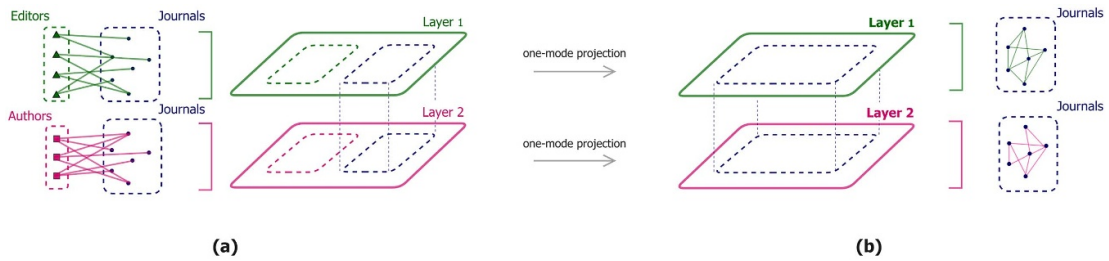

Figure 1. Scheme depicting an example of multilayer network of journals, editors and authors with two layers, each containing a bipartite graph. We start by considering a multilayer network with a bipartite network on each layer (a). In (a), the set of journals (blue nodes) is common to all the layers, while the green and pink sets represent editors and authors, respectively. The one mode projection on the set of journals allows to obtain the multiplex network (b), where nodes on each layer are journals.

expression (2). In addition, Lunagómez *et al* (2021) consider generalizations of the model by Banks and Carley (1994) and emphasize the central role of the sample Fréchet mean for the location parameter estimation under the Bayesian paradigm. Hence, there is an interesting connection between our descriptive approach and the estimation of the central network under the model-based approach.

In the following subsections, we discuss the adoption of three appropriate metrics d, i.e. the Frobenius, the Riemannian and the Wasserstein metrics, and the properties of the corresponding $\boldsymbol{S}_+$.

4.1. Frobenius metric

If the classical Frobenius metric is adopted, i.e.

$$d_{\mathrm{F}}(\boldsymbol{X},\boldsymbol{Y}) = \|\boldsymbol{X}-\boldsymbol{Y}\|_{\mathrm{F}} = \mathrm{tr}((\boldsymbol{X}-\boldsymbol{Y})^{\mathrm{T}}(\boldsymbol{X}-\boldsymbol{Y}))^{1/2}$$

for $\boldsymbol{X},\boldsymbol{Y}\in\mathcal{P}_n$, the weighted mean suggested in Abdi *et al* (2005) is obtained

$$\boldsymbol{S}_{+\mathrm{F}} = \sum_{k=1}^{m} w_k \boldsymbol{S}_k. \tag{3}$$

From (*i*) of proposition 2.1, $\boldsymbol{S}_{+\mathrm{F}}\in\mathcal{CP}_n$ as expected, since $\boldsymbol{S}_{+\mathrm{F}}$ is a weighted sum of completely positive matrices with positive weights. However, this proposal may involve a 'swelling effect' as a drawback, in the sense that $\boldsymbol{S}_{+\mathrm{F}}$ may show an increase in the determinant with respect to the single components of the mean (see e.g. Fletcher and Joshi 2007). To be more precise, if a given matrix average $\boldsymbol{S}_+$ is considered, the swelling effect generally occurs when $\det(\boldsymbol{S}_+)\geqslant\det(\boldsymbol{S}_k)$ for each k (Arsigny *et al* 2007).

4.2. Riemannian metric

Let us assume the Riemannian metric, i.e.

$$\begin{aligned} d_{\mathrm{R}}(\boldsymbol{X},\boldsymbol{Y}) &= \|\log(\boldsymbol{X}^{-1/2}\boldsymbol{Y}\boldsymbol{X}^{-1/2})\|_{\mathrm{F}} \\ &= \mathrm{tr}(\log(\boldsymbol{X}^{-1/2}\boldsymbol{Y}\boldsymbol{X}^{-1/2})^{\mathrm{T}}\log(\boldsymbol{X}^{-1/2}\boldsymbol{Y}\boldsymbol{X}^{-1/2}))^{1/2} \end{aligned}$$

for $\boldsymbol{X},\boldsymbol{Y}\in\mathcal{P}_n$. Bhatia (2009, chapter 6, theorem 6.1.6) remarks that d_{R} naturally arises in the framework of Riemannian geometry (for further details, see also Bhatia and Congedo 2019, and references therein). Moreover, d_{R} may be considered as the matrix version of the Fisher-Rao metric for probability laws (see in turn chapter 6 in Bhatia 2009). In this case, the matrix $\boldsymbol{S}_{+\mathrm{R}}$ exists and is the unique solution of the nonlinear matrix equation in $\boldsymbol{X}$ given by

$$\sum_{k=1}^{m} w_k \log(\boldsymbol{X}^{1/2}\boldsymbol{S}_k^{-1}\boldsymbol{X}^{1/2}) = \boldsymbol{0}, \tag{4}$$

even if it is not generally expressible in closed form (see e.g. Lim and Pálfia 2014a, 2014b).

In order to introduce the explicit expression of $\boldsymbol{S}_{+\mathrm{R}}$ for $m=2$, let us consider the weighted geometric mean of two matrices $\boldsymbol{X},\boldsymbol{Y}\in\mathcal{P}_n$, i.e.

$$\boldsymbol{X}\#_w\boldsymbol{Y} = \boldsymbol{X}(\boldsymbol{X}^{-1}\boldsymbol{Y})^w = \boldsymbol{X}^{1/2}(\boldsymbol{X}^{-1/2}\boldsymbol{Y}\boldsymbol{X}^{-1/2})^w\boldsymbol{X}^{1/2}, \tag{5}$$

where $w \in [0,1]$ (for its properties, see Bhatia 2009, and for its efficient computation see Iannazzo 2016). Obviously, if $\boldsymbol{X}$ and $\boldsymbol{Y}$ were scalars and $w = 1/2$, expression (5) reduces to the usual geometric mean of two numbers. Moreover, if $\boldsymbol{X}, \boldsymbol{Y} \in \mathcal{P}_n$ it follows that $\boldsymbol{X} \#_w \boldsymbol{Y} \in \mathcal{P}_n$ (Lim and Pálfia 2014a). Actually, for $m = 2$ it can be proven that the solution of the nonlinear matrix equation (4) is given by

$$\boldsymbol{S}_{+\mathrm{R}} = \boldsymbol{S}_2 \#_{w_1} \boldsymbol{S}_1 = \boldsymbol{S}_2^{1/2} (\boldsymbol{S}_2^{-1/2} \boldsymbol{S}_1 \boldsymbol{S}_2^{-1/2})^{w_1} \boldsymbol{S}_2^{1/2} \tag{6}$$

(see e.g. Bhatia 2009, p 210). This is the reason for which $\boldsymbol{S}_{+\mathrm{R}}$ is named as the weighted geometric mean of positive definite matrices also for a general m. It should be remarked that $\boldsymbol{S}_{+\mathrm{R}} = \boldsymbol{S}_1$ if $w_1 = 1$ and $\boldsymbol{S}_{+\mathrm{R}} = \boldsymbol{S}_2$ if $w_1 = 0$. Moreover, since $\boldsymbol{S}_{+\mathrm{R}}$ constitutes a geodesic from $\boldsymbol{S}_1$ to $\boldsymbol{S}_2$ for varying $w_1 \in [0,1]$ (see e.g. Bhatia 2009, p 210) and $\mathcal{CP}_n$ is a closed convex cone in $\mathcal{P}_n$, then $\boldsymbol{S}_{+\mathrm{R}} \in \mathcal{CP}_n$ for $m = 2$.

If $m \geqslant 3$, Lim and Pálfia (2014a) propose an iterative procedure for computing $\boldsymbol{S}_{+\mathrm{R}}$. Let us define the iterative sequence

$$\boldsymbol{X}_{\ell+1} = \boldsymbol{X}_\ell \#_{t_{\ell+1}} \boldsymbol{S}_{j_{\ell+1}} \tag{7}$$

for $\ell = 1, 2, \ldots$, where $\boldsymbol{X}_1 = \boldsymbol{S}_{j_1}$ and $t_\ell = w_{j_\ell} / \sum_{i=1}^{\ell} w_{j_i}$. In addition, $j_\ell = (\ell \bmod m)$ and null residuals are identified with m, i.e. $j_{im} = m$ for $i = 1, 2, \ldots$. Accordingly, Lim and Pálfia (2014a) prove that $\lim_\ell \boldsymbol{X}_\ell = \boldsymbol{S}_{+\mathrm{R}}$. In Massart *et al* (2018), an alternative choice leading to a more efficient computation of the index j_k is provided. Since at each step $\boldsymbol{X}_\ell \in \mathcal{CP}_n$, it also follows that $\boldsymbol{S}_{+\mathrm{R}} \in \mathcal{CP}_n$. Moreover, $\boldsymbol{S}_{+\mathrm{R}} \preceq \boldsymbol{S}_{+\mathrm{F}}$ holds on the basis of the generalized arithmetic-geometric-harmonic mean inequality (see Lim and Pálfia 2014b), a result which emphasizes that $\boldsymbol{S}_{+\mathrm{R}}$ may be less prone than $\boldsymbol{S}_{+\mathrm{F}}$ to the swelling effect.

4.3. Wasserstein metric

A further proposal for d is given by the Wasserstein metric, i.e.

$$d_{\mathrm{W}}(\boldsymbol{X}, \boldsymbol{Y}) = \mathrm{tr}(\boldsymbol{X} + \boldsymbol{Y} - 2(\boldsymbol{X}^{1/2} \boldsymbol{Y} \boldsymbol{X}^{1/2})^{1/2})^{1/2}$$

for $\boldsymbol{X}, \boldsymbol{Y} \in \mathcal{P}_n$ (for more details, see e.g. Bhatia *et al* 2019). Bhatia *et al* (2019) emphasize that d_{W} displays many interesting features and, among others, it corresponds to a metric in Riemannian geometry. Álvarez-Esteban *et al* (2016) prove that $\boldsymbol{S}_{+\mathrm{W}}$ is provided by the unique solution of the nonlinear matrix equation in $\boldsymbol{X}$ given by

$$\boldsymbol{X} = \sum_{k=1}^{m} w_k (\boldsymbol{X}^{1/2} \boldsymbol{S}_k \boldsymbol{X}^{1/2})^{1/2} \tag{8}$$

and $\boldsymbol{S}_{+\mathrm{W}} \in \mathcal{P}_n$ (see also Bhatia *et al* 2019). The solution of the nonlinear matrix equation (8) is solely known in a closed form for $m = 2$ and it reads

$$\boldsymbol{S}_{+\mathrm{W}} = w_1^2 \boldsymbol{S}_1 + w_2^2 \boldsymbol{S}_2 + w_1 w_2 ((\boldsymbol{S}_1 \boldsymbol{S}_2)^{1/2} + (\boldsymbol{S}_2 \boldsymbol{S}_1)^{1/2}) \tag{9}$$

(see Bhatia *et al* 2019). Incidentally, it is interesting to remark that, if $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ were scalars and $w_1 = w_2 = 1/2$, expression (9) actually provides the average of the usual arithmetic and geometric means of two numbers. Similarly to the case of the weighted geometric mean, it holds that $\boldsymbol{S}_{+\mathrm{W}} = \boldsymbol{S}_1$ if $w_1 = 1$ and $\boldsymbol{S}_{+\mathrm{R}} = \boldsymbol{S}_2$ if $w_1 = 0$, while $\boldsymbol{S}_{+\mathrm{W}}$ is a geodesic from $\boldsymbol{S}_1$ to $\boldsymbol{S}_2$ for varying $w_1 \in [0,1]$ (Bhatia *et al* 2019) and hence $\boldsymbol{S}_{+\mathrm{R}} \in \mathcal{CP}_n$ for $m = 2$.

In order to manage the case $m \geqslant 3$ by adopting an algorithm based on a fixed-point iteration method, Álvarez-Esteban *et al* (2016) suggest to consider the matrix function $\boldsymbol{K} : \mathcal{P}_n \to \mathcal{P}_n$ such that

$$\boldsymbol{K}(\boldsymbol{X}) = \boldsymbol{X}^{-1/2} \left(\sum_{i=1}^{m} w_i (\boldsymbol{X}^{1/2} \boldsymbol{S}_i \boldsymbol{X}^{1/2})^{1/2} \right)^2 \boldsymbol{X}^{-1/2}. \tag{10}$$

Álvarez-Esteban *et al* (2016) highlight that $\boldsymbol{K}(\boldsymbol{X}) \in \mathcal{P}_n$ if $\boldsymbol{X} \in \mathcal{P}_n$. Hence, by assuming that $\boldsymbol{X}_{\ell+1} = \boldsymbol{K}(\boldsymbol{X}_\ell)$ for $\ell = 0, 1, \ldots$ and for each $\boldsymbol{X}_0 \in \mathcal{P}_n$, Álvarez-Esteban *et al* (2016) prove that $\lim_\ell \boldsymbol{X}_\ell = \boldsymbol{S}_{+\mathrm{W}}$. In addition, by means of a numerical study, they remark that algorithm convergence is fast, even for rather large n and m. It is worth noticing that $\boldsymbol{S}_{+\mathrm{W}} \preceq \boldsymbol{S}_{+\mathrm{F}}$ on the basis of theorem 9 by Bhatia (2009), which is a suitable property, as previously explained.

4.4. Pros and cons of different metric choices

In this section we illustrate the advantages and drawbacks brought by the choices of the considered metrics. Concerning the Frobenius metric, this proposal produces an average similarity matrix $\boldsymbol{S}_{+\mathrm{F}}$ which is easily interpreted from a statistical perspective, since it is a weighted sum of the similarity matrices $\boldsymbol{S}_1, \ldots, \boldsymbol{S}_m$. For this reason, $\boldsymbol{S}_{+\mathrm{F}}$ is straightforwardly computed without requiring ad-hoc algorithms, even for large n and m. Moreover, $\boldsymbol{S}_{+\mathrm{F}}$ is a bona-fide normalized similarity matrix. However, this metric selection may lead to the swelling effect—a phenomenon which, in data analysis, may generally add harmful spurious variations to the data (see e.g. the discussion in the signal processing setting provided by Yger *et al* 2017). In network analysis, the swelling effect produces an increase in the determinant of $\boldsymbol{S}_{+\mathrm{F}}$ and, consequently, an increase in the corresponding eigenvalues (which are positive on the basis of the assumptions). This issue could lead to a growth of the centrality of the single nodes in the average network, as the centrality has a connection with the eigenvalues of the similarity matrix (see Newman 2018, section 7.1).

As emphasized in sections 4.2 and 4.3, the metrics d_{R} and d_{W} are very appropriate for the space $\mathcal{P}_n$, and hence for the space $\mathcal{CP}_n$. Moreover, the use of d_{R} and d_{W} involves an optimal rotation of each couple of similarity matrices, i.e. they actually provide the minimum for the orthogonal Procrustes problem (see comment to theorem 1 in Bhatia 2009, and Bhatia and Congedo 2019). In addition, the swelling effect is reduced with respect to the use of d_{F} (it follows from the results given in sections 4.2 and 4.3). However, the computation of $\boldsymbol{S}_{+\mathrm{R}}$ and $\boldsymbol{S}_{+\mathrm{W}}$ is involved for $m \geqslant 3$ and requires specific iterative algorithms, which could be slow and could have a loss of precision for large n. In the case of $\boldsymbol{S}_{+\mathrm{R}}$, an efficient algorithm for the computation of the weighted geometric means of two matrices is in turn necessary. We finally remark that $\boldsymbol{S}_{+\mathrm{R}}$ and $\boldsymbol{S}_{+\mathrm{W}}$ have to be re-normalized in order to achieve a proper normalized similarity matrix having all values equal to 1 on the diagonal. In such a case, a suitable way to achieve a normalized similarity matrix $\boldsymbol{S}^* = (s^*_{ij})$ consists in modifying the generic element s_{ij} of a similarity matrix $\boldsymbol{S} = (s_{ij})$ as follows

$$s^*_{ij} = \frac{s_{ij}}{(s_{ii}s_{jj})^{1/2}}.$$

It is easily verified that if $\boldsymbol{S} \in \mathcal{CP}_n$ also $\boldsymbol{S}^* \in \mathcal{CP}_n$.

5. Choice of the weights

In order to select the weights $\boldsymbol{w} = (w_1, \ldots, w_m)^{\mathrm{T}}$, a measure of 'closeness' between the couples of the m similarity matrices $\boldsymbol{S}_1, \ldots, \boldsymbol{S}_m$ is needed at first. A suitable such a measure is given by the RV coefficient proposed by Robert and Escoufier (1976), which is widely adopted in many different practical analyses. Hence, the matrix $\boldsymbol{R} = (r_{ij})$ of order $(m \times m)$ may be considered, where r_{ij} represents the RV coefficient between $\boldsymbol{S}_i$ and $\boldsymbol{S}_j$, i.e.

$$r_{ij} = \frac{\langle \boldsymbol{S}_i, \boldsymbol{S}_j \rangle_{\mathrm{F}}}{\|\boldsymbol{S}_i\|_{\mathrm{F}} \|\boldsymbol{S}_j\|_{\mathrm{F}}}.$$

It holds that $r_{ij} \in [0, 1]$ and the closeness between $\boldsymbol{S}_i$ and $\boldsymbol{S}_j$ increases as r_{ij} approaches one. In addition, it should be remarked that $\boldsymbol{R} \in \mathcal{CP}_m$. Indeed, this issue follows from (*iii*) of proposition 2.1, since $\boldsymbol{R}$ may be expressed as $\boldsymbol{R} = \mathrm{diag}(\boldsymbol{U})^{-1/2}\boldsymbol{U}\mathrm{diag}(\boldsymbol{U})^{-1/2}$ where $\boldsymbol{U} = \boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}$ and $\boldsymbol{V} = (\mathrm{vec}(\boldsymbol{S}_1), \ldots, \mathrm{vec}(\boldsymbol{S}_m))$, which also provides a computationally efficient expression for $\boldsymbol{R}$.

When $\boldsymbol{S}_{+\mathrm{F}}$ is adopted, Abdi *et al* (2005) suggest to consider the eigendecomposition of $\boldsymbol{R}$. In such a case, the decomposition gives rise to $\boldsymbol{R} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^{\mathrm{T}}$, where $\boldsymbol{Q} = (\boldsymbol{q}_1, \ldots, \boldsymbol{q}_m)$ is the orthogonal matrix whose columns are the eigenvectors of $\boldsymbol{R}$, while $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1(\boldsymbol{R}), \ldots, \lambda_m(\boldsymbol{R}))$ is the diagonal matrix whose entries are the positive eigenvalues of $\boldsymbol{R}$ considered in nonincreasing order. The Perron–Frobenius theorem (see e.g. Berman and Plemmons 1994) ensures that the eigenvector $\boldsymbol{q}_1$ corresponding to the largest eigenvalue $\lambda_1(\boldsymbol{R})$ has nonnegative elements. Thus, Abdi *et al* (2005) propose the choice

$$\boldsymbol{w} = \frac{1}{\boldsymbol{1}^{\mathrm{T}}\boldsymbol{q}_1}\boldsymbol{q}_1.$$

In practice, a principal component analysis is considered on $\boldsymbol{R}$ and the first eigenvector is used for implementing $\boldsymbol{w}$. Hence, since layers with larger projections on $\boldsymbol{q}_1$ are more similar to the other layers than the layers with smaller projections, the (rescaled) elements of this eigenvector should provide suitable weights for $\boldsymbol{S}_{+\mathrm{F}}$, which is a linear combination of $\boldsymbol{S}_1, \ldots, \boldsymbol{S}_m$. Similarly to principal component analysis, the goodness of this selection for $\boldsymbol{w}$ may be assessed by means of the ratio $\lambda_1(\boldsymbol{R}) / \sum_{k=1}^{m} \lambda_k(\boldsymbol{R})$ (see Abdi *et al* 2005).

In the case of $\boldsymbol{S}_{+\mathrm{R}}$ and $\boldsymbol{S}_{+\mathrm{W}}$ it is not obvious if the previous choice of $\boldsymbol{w}$ is suitable, since these averages are not linear functions of $\boldsymbol{S}_1, \ldots, \boldsymbol{S}_m$. Alternatively, we suggest the choice

$$\boldsymbol{w} = \frac{1}{\mathbf{1}^{\mathrm{T}}(\boldsymbol{R} - \boldsymbol{I}_m)\mathbf{1}}(\boldsymbol{R} - \boldsymbol{I}_m)\mathbf{1}.$$

The rationale behind this proposal stems on the fact that the more a similarity matrix is close to the others, the more it is representative of the whole set of similarity matrices—and hence it should receive a larger weight with respect to the others. Thus, this proposal assigns weights according to the ratio of the RV coefficient sum corresponding to a similarity matrix to the total of the RV coefficients. Hence, this choice could be generally suitable, while the weights proposed by Abdi *et al* (2005) are more appropriate when the Frobenius metric is specifically considered.

6. Implementation of algorithms for computing averages

The algorithms for the computation of the similarity matrix averages $\boldsymbol{S}_{+\mathrm{F}}$, $\boldsymbol{S}_{+\mathrm{W}}$ and $\boldsymbol{S}_{+\mathrm{R}}$ were implemented using the Python programming language. All the Python functions are publicly available on GitHub at https://github.com/DedeBac/SimilarityMatrixAggregation.git.

The choice of the Frobenius metric leads to a similarity matrix average which is actually a weighted arithmetic mean of matrices—see expression (3). Hence, the practical implementation of $\boldsymbol{S}_{+\mathrm{F}}$ is trivial. In contrast, as discussed in sections 4.2 and 4.3, the choice of the Riemannian and the Wasserstein metrics requires that $\boldsymbol{S}_{+\mathrm{R}}$ and $\boldsymbol{S}_{+\mathrm{W}}$ are computed via iterative algorithms, unless there are solely two input matrices. Thus, in the special case $m = 2$ (discussed in sections 4.2 and 4.3), the closed-form solutions available in expressions (6) and (9) were implemented. In the non-trivial case $m > 2$, $\boldsymbol{S}_{+\mathrm{R}}$ is computed using the iterative sequence defined in expression (7). In turn, such an implementation is based on the optimized procedure proposed by Massart *et al* (2018), as introduced in section 4.2. A key issue of the iterative step consists in the computation of the weighted geometric mean of two matrices. In our algorithm, the weighted geometric mean was implemented by using the proposal by Iannazzo (2016) (algorithm 3.1). Specifically, this implementation adopts the Schur decomposition and the Cholesky factorization in order to simplify the computation of the power of a matrix. In the case of the Wasserstein metric, $\boldsymbol{S}_{+\mathrm{W}}$ is computed by means of the fixed-point iteration method proposed by Álvarez-Esteban *et al* (2016) and discussed in section 4.3, see equation (10). Both these iterative procedures stop when either a maximum number of iterations is reached, or when the relative distance between intermediate solutions computed at consecutive iterations reaches the tolerance value set by the user. Finally, the implementation of the weights computation follows the proposals discussed in section 5.

We provide a visual example of the computation of the matrix averages for each metric choice in the very special case $n = 2$ and $m = 2$. In such a setting, the results given in the previous sections may be conveniently—and instructively—visualized by means of ellipses of type $\{\boldsymbol{x} \in \mathbb{R}^2 : \boldsymbol{x}^{\mathrm{T}}\boldsymbol{S}\boldsymbol{x} = 1\}$. Thus, for $w_1 = w_2 = 1/2$, figures 2, 3 and 4 provide the graphical representation of $\boldsymbol{S}_{+\mathrm{F}}$, $\boldsymbol{S}_{+\mathrm{R}}$ and $\boldsymbol{S}_{+\mathrm{W}}$ for several choices of $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$. From these figures, it is apparent that $\boldsymbol{S}_{+\mathrm{R}}$ and $\boldsymbol{S}_{+\mathrm{W}}$ are less affected by the swelling effect (see especially figure 4). Moreover, even if $\boldsymbol{S}_{+\mathrm{F}}$, $\boldsymbol{S}_{+\mathrm{R}}$ and $\boldsymbol{S}_{+\mathrm{W}}$ have rather similar 'shapes' in figures 2 and 3, these 'shapes' substantially differ in figure 4.

7. Case study: multiplex network of statistical journals

In this section we present an application of the use of matrix averages to multiplex network of statistical journals. Specifically, we will consider 79 journals belonging to the category 'Statistics and Probability' of the *Journal Citation Reports* in the year 2006. Each journal was associated with three different types of 'items' in order to obtain the multiplex network: the editors sitting on the editorial boards, the authors that published on the journals and the journals cited in the bibliographies of the corresponding articles. This kind of connections have a key relevance in a scientometric setting (see Baccini *et al* 2020, 2022, and references therein).

Editor data include the editors of the journals in the year 2006 and were collected by Baccini *et al* (2009). On the other hand, data on authors and cited journals were extracted from Clarivate Analytics Web of Science (WoS) database (https://www.webofscience.com/). Queries were made through the SQL relational database system hosted by the Centre for Science and Technology Studies (CWTS) at Leiden University,

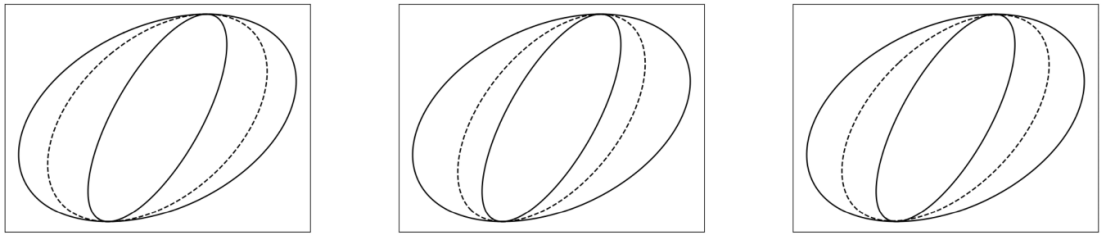

Figure 2. Visualization of (a) $\boldsymbol{S}_{+\mathrm{F}}$, (b) $\boldsymbol{S}_{+\mathrm{R}}$ and (c) $\boldsymbol{S}_{+\mathrm{W}}$ (dashed style) for $\boldsymbol{S}_1 = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$ and $\boldsymbol{S}_2 = \begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix}$. Continuous lines correspond to the ellipses individuated by $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$.

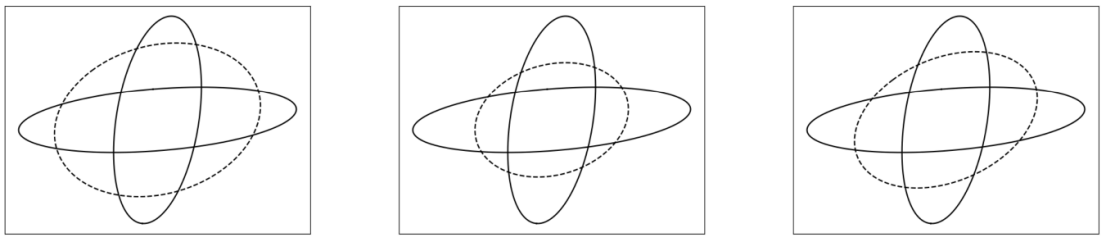

Figure 3. Visualization of (a) $\boldsymbol{S}_{+\mathrm{F}}$, (b) $\boldsymbol{S}_{+\mathrm{R}}$ and (c) $\boldsymbol{S}_{+\mathrm{W}}$ (dashed style) for $\boldsymbol{S}_1 = \begin{pmatrix} 1 & 1 \\ 1 & 10 \end{pmatrix}$ and $\boldsymbol{S}_2 = \begin{pmatrix} 10 & 1 \\ 1 & 1 \end{pmatrix}$. Continuous lines correspond to the ellipses individuated by $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$.

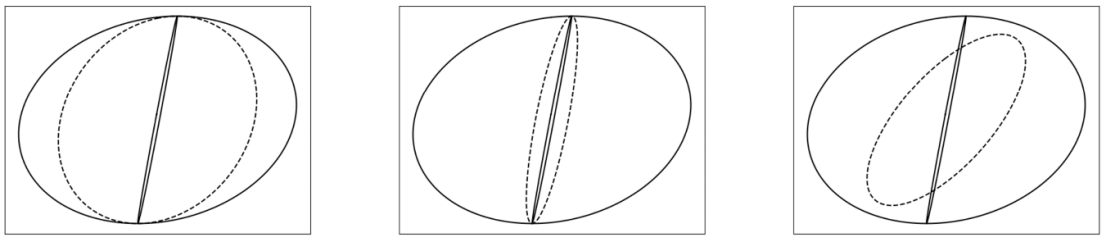

Figure 4. Visualization of (a) $\boldsymbol{S}_{+\mathrm{F}}$, (b) $\boldsymbol{S}_{+\mathrm{R}}$ and (c) $\boldsymbol{S}_{+\mathrm{W}}$ (dashed style) for $\boldsymbol{S}_1 = \begin{pmatrix} 50 & 1 \\ 1 & 1 \end{pmatrix}$ and $\boldsymbol{S}_2 = \begin{pmatrix} 1.01 & 1 \\ 1 & 1 \end{pmatrix}$. Continuous lines correspond to the ellipses individuated by $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$.

using the 2021 version of Web of Science. In both cases, authorships and cited journals were solely collected for publications in the five-year period 2006–2010. Table 1 offers an overview of the considered dataset.

Before proceeding to the computation of averages, the journal *Communications in Statistics—Simulation and Computation* was removed from the dataset since it had exactly the same editors of the journal *Communications in Statistics—Theory and Methods*. In fact, Taylor & Francis, the publisher of the two journals, considers them as 'associated journals'. Thus, in order to avoid ambiguity we solely retained *Communications in Statistics—Theory and Methods*.

The resulting one-mode networks respectively model the pairwise connections between journals based on common editors (interlocking editorship network), common authors (interlocking authorship network) and common cited journals (journal bibliographic coupling network). To the aim of computing the analyzed averages, the cosine similarity defined in equation (1) was considered for each layer in order to obtain three

Table 1. Main features of the considered scientometric dataset.

Layer	Period of observation	Two-mode network	Relation in the one-mode network
Editors	2006	79 journals × 2227 editors	Interlocking editorship
Authors	2006–2010	79 journals × 38 683 authors	Interlocking authorship
Cited journals	2006–2010	79 journals × 7528 cited journals	Journal bibliographic coupling

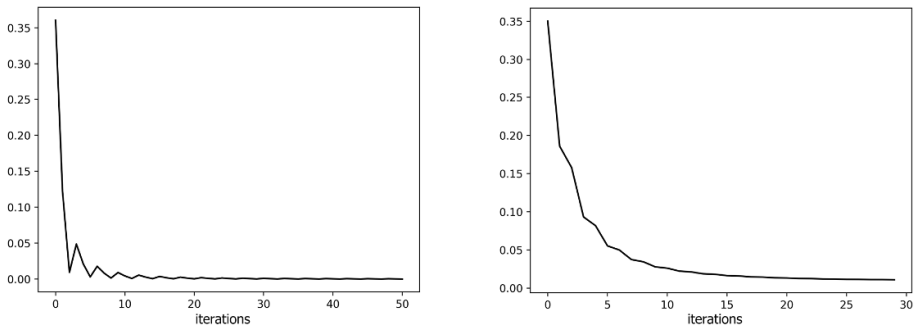

Figure 5. Residual values of the algorithm iterative steps for the computation of $\boldsymbol{S}_{+\mathrm{W}}$ for 50 iterations (a) and $\boldsymbol{S}_{+\mathrm{W}}$ for 30 iterations (b).

similarity matrices. The averages $\boldsymbol{S}_{+\mathrm{F}}$, $\boldsymbol{S}_{+\mathrm{R}}$ and $\boldsymbol{S}_{+\mathrm{W}}$ of the three similarity matrices were computed using the implementation described in section 6. Figure 5 displays the residual between matrices computed at consecutive iterations in the case of $\boldsymbol{S}_{+\mathrm{R}}$ (figure 5(a)) and $\boldsymbol{S}_{+\mathrm{W}}$ (figure 5(b)). It should be remarked that both algorithms converge to a solution after few iterative steps.

Figures 6–8 respectively display the networks corresponding to the averages $\boldsymbol{S}_{+\mathrm{F}}$, $\boldsymbol{S}_{+\mathrm{R}}$ and $\boldsymbol{S}_{+\mathrm{W}}$. The networks were plotted by means of the Gephi software (Bastian *et al* 2009) and by using the ForceAtlas2 visualization algorithm (Jacomy *et al* 2014). Different colors correspond to distinct communities computed with the Louvain algorithm (Blondel *et al* 2008) by setting the resolution parameter to 1 (see Lambiotte *et al* 2008). Actually, the Louvain algorithm is commonly adopted in order to detect communities in a network. The modularity score obtained was 0.131 for $\boldsymbol{S}_{+\mathrm{F}}$, 0.222 for $\boldsymbol{S}_{+\mathrm{R}}$, and 0.166 for $\boldsymbol{S}_{+\mathrm{W}}$, as reported in table 2. The Louvain algorithm partitions the three aggregated networks into five communities. All the networks are characterized by three large communities, indicated by the pink, orange and green nodes in figures 6–8, and reported in table 2 as Community 1, Community 2 and Community 3, respectively. Specifically, Community 1 (pink nodes) gathers journals in the field of statistical methodology, Community 2 (orange nodes) groups journals mainly devoted to probability theory and its applications, while Community 3 gathers journals from the field of applied statistics. Finally, two much smaller communities are individuated—Community 4 (blue nodes) and Community 5 (yellow nodes) in table 2. They both gather journals that are at the boundary with the fields of Economics (Community 4) and Bioinformatics (Community 5), respectively.

Overall, the modularity values and the communities detected in the three aggregated networks are very similar. This issue suggests that the choice of the metric in the computation of the matrix average does not significantly affect the community structure of the network. Moreover, the emerging communities are coherent with those obtained in Baccini *et al* (2020, 2022).

Finally, in order to assess the structural similarity between $\boldsymbol{S}_{+\mathrm{F}}$, $\boldsymbol{S}_{+\mathrm{R}}$, and $\boldsymbol{S}_{+\mathrm{W}}$, the generalized distance correlation suggested by Székely *et al* (2007) is computed for each pair of networks. Distance correlation is a measure of correlation between distance matrices belonging to the family of 'energy statistics' (Székely and Rizzo 2017, Szekely and Rizzo 2023). Such measure can be seen as a generalization to matrix spaces of the usual squared Pearson correlation coefficient. Indeed, it assumes values in the interval $[0, 1]$, with values near to zero indicating a low level of association, and values close to one indicating high association. Omelka and Hudecová (2013) observed that its use is suitable to evaluate the association between similarity matrices. The distance correlation values between the similarity matrices $\boldsymbol{S}_{+\mathrm{F}}$, $\boldsymbol{S}_{+\mathrm{R}}$, and $\boldsymbol{S}_{+\mathrm{W}}$ are reported in table 3. The table shows that a very elevate association exists between each pair of averages. Therefore, the monoplex networks obtained with different metric choices are strongly associated, i.e. they are structurally similar.

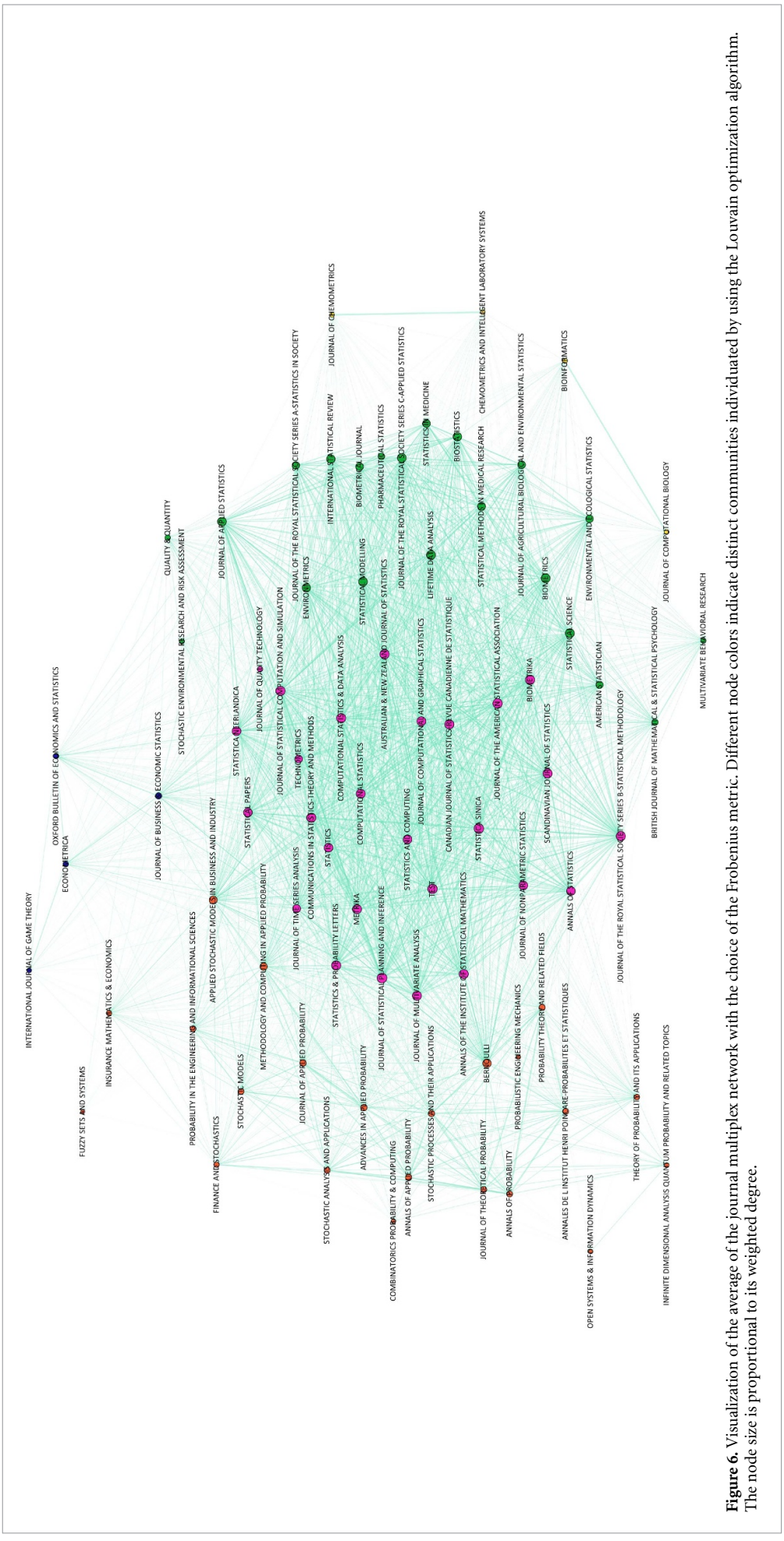

Figure 6. Visualization of the average of the journal multiplex network with the choice of the Frobenius metric. Different node colors indicate distinct communities individuated by using the Louvain optimization algorithm. The node size is proportional to its weighted degree.

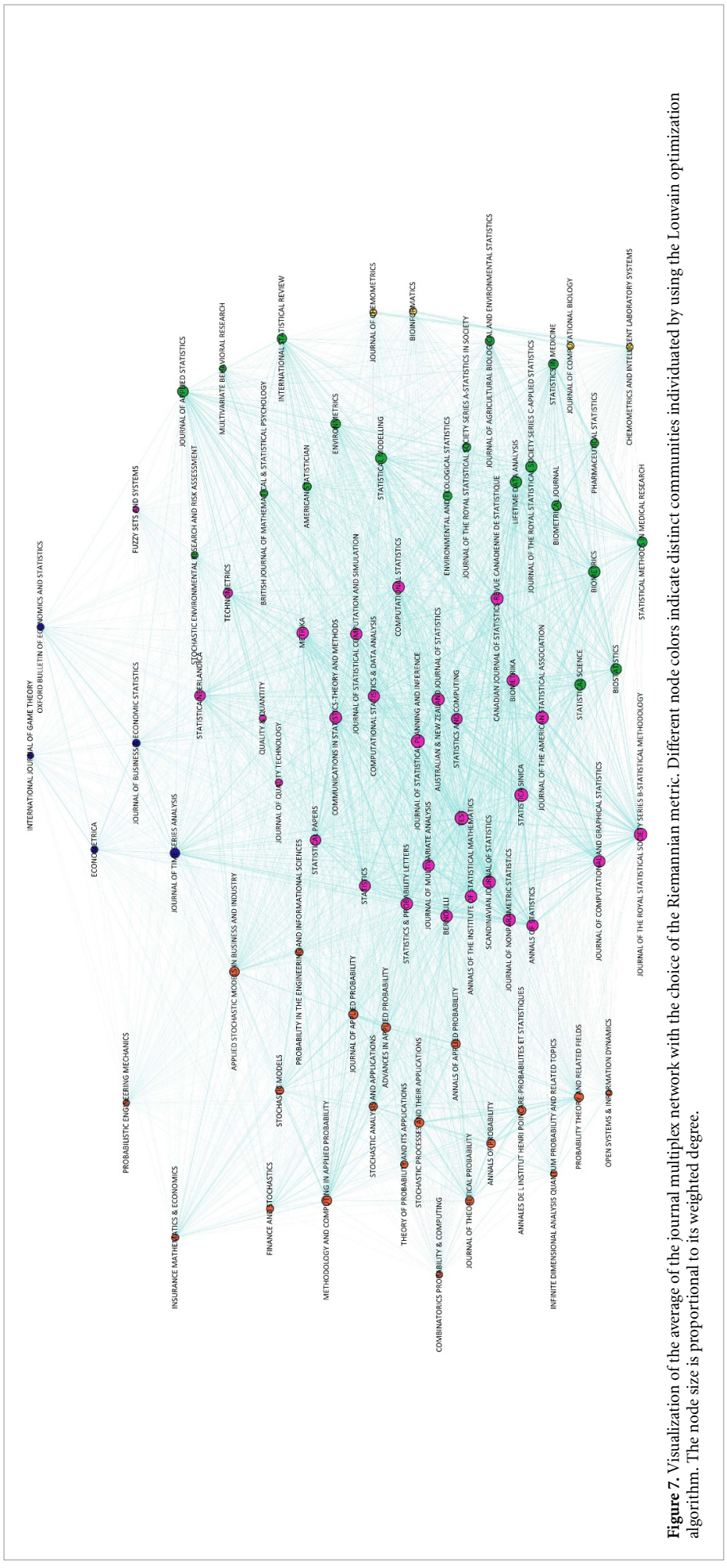

Figure 7. Visualization of the average of the journal multiplex network with the choice of the Riemannian metric. Different node colors indicate distinct communities individuated by using the Louvain optimization algorithm. The node size is proportional to its weighted degree.

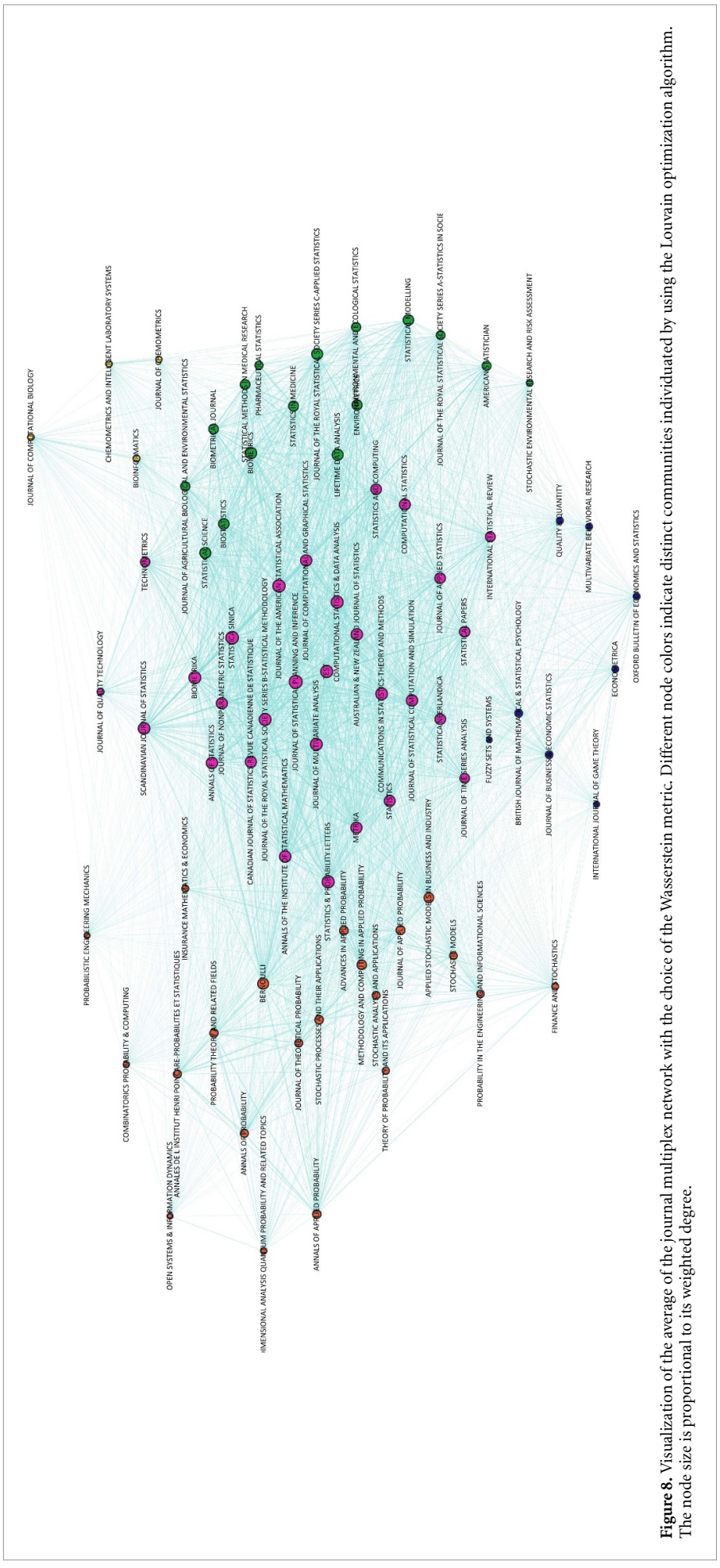

Figure 8. Visualization of the average of the journal multiplex network with the choice of the Wasserstein metric. Different node colors indicate distinct communities individuated by using the Louvain optimization algorithm. The node size is proportional to its weighted degree.

Table 2. Number of nodes for each community individuated by the Louvain modularity optimization algorithm and for each average $\boldsymbol{S}_{+F}$, $\boldsymbol{S}_{+R}$ and $\boldsymbol{S}_{+W}$. The first column reports the modularity score associated to each partitioning of the networks.

Average	Modularity	Community 1	Community 2	Community 3	Community 4	Community 5
$\boldsymbol{S}_{+F}$	0.131	27	22	21	4	4
$\boldsymbol{S}_{+R}$	0.222	29	20	20	5	4
$\boldsymbol{S}_{+W}$	0.166	29	21	16	8	4

Table 3. Distance correlation between monoplex networks obtained with the choice of the Frobenius, Riemannian and Wasserstein metrics (denoted, respectively, as $\boldsymbol{S}_{+F}$, $\boldsymbol{S}_{+R}$, and $\boldsymbol{S}_{+W}$). Since distance correlation is symmetric in its arguments, we only fill the upper triangular part of the table.

	$\boldsymbol{S}_{+F}$	$\boldsymbol{S}_{+R}$	$\boldsymbol{S}_{+W}$
$\boldsymbol{S}_{+F}$	1	0.9634	0.9885
$\boldsymbol{S}_{+R}$	—	1	0.9896
$\boldsymbol{S}_{+W}$	—	—	1

8. Conclusions

This work proposes the use of similarity matrix average to aggregate multiplex networks by using some concepts in the Riemannian geometry formulation of barycenter. In particular, we show that some commonly-adopted similarity measures allow to obtain similarity matrices which belong to the space of completely positive matrices by starting from two-mode bipartite networks. These types of similarity matrices are averaged by considering the minimization of a Fréchet mean criterion with the Frobenius, the Riemannian and the Wasserstein metric choices. The results obtained on the multiplex network of statistical journals highlight the availability of the methodology. The proposed method constitutes an advance in the topic of multiplex network aggregation, since it provides a theoretically justified framework—along with its implementation—to combine the contribution of different relations which exist among a set of nodes. This methodology might be of interest in several application fields, such as social network analysis and bioinformatics, where relations of different nature have to be explored and integrated to determine some structural organization of a set of entities. Indeed, the aggregation allows to perform the required task—such as the detection of communities—on a single network, rather than on a multiplex network, in order to reduce the complexity of the problem.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://github.com/DedeBac/SimilarityMatrixAggregation.git.

Acknowledgments

Eugenio Petrovich is grateful to the Centre for Science and Technology Studies (CWTS) at Leiden University for hosting him as a guest researcher and giving him access to the CWTS database system. Lucio Barabesi gratefully acknowledges the funding from the Italian Ministry of University, PRIN Project: 2017MPXW98. All the authors would like to acknowledge the Editor and the Reviewers for their useful comments.

ORCID iD

Federica Baccini https://orcid.org/0000-0001-6107-9611

References

Abdi H, O'Toole A J, Valentin D and Edelman B 2005 DISTATIS: the analysis of multiple distance matrices *Proc. IEEE Computer Society: Int. Conf. on Computer Vision and Pattern Recognition (San Diego, CA)* pp 42–47

Álvarez-Esteban P C, Del Barrio E, Cuesta-Albertos J and Matrán C 2016 A fixed-point approach to barycenters in Wasserstein space *J. Math. Anal. Appl.* **441** 744–62

Argelaguet R, Arnol D, Bredikhin D, Deloro Y, Velten B, Marioni J C and Stegle O 2020 Mofa+: a statistical framework for comprehensive integration of multi-modal single-cell data *Genome Biol.* **21** 1–17

Argelaguet R, Velten B, Arnol D, Dietrich S, Zenz T, Marioni J C, Buettner F, Huber W and Stegle O 2018 Multi-omics factor analysis—a framework for unsupervised integration of multi-omics data sets *Mol. Syst. Biol.* **14** e8124

Arsigny V, Fillard P, Pennec X and Ayache N 2007 Geometric means in a novel vector space structure on symmetric positive-definite matrices *SIAM J. Matrix Anal. Appl.* **29** 328–47

Bacák M 2014 Computing medians and means in Hadamard spaces *SIAM J. Optim.* **24** 1542–66
Baccini A, Barabesi L, Khelfaoui M and Gingras Y 2020 Intellectual and social similarity among scholarly journals: an exploratory comparison of the networks of editors, authors and co-citations *Quant. Sci. Stud.* **1** 277–89
Baccini A, Barabesi L and Marcheselli M 2009 How are statistical journals linked? A network analysis *CHANCE* **22** 35–45
Baccini F, Barabesi L, Baccini A, Khelfaoui M and Gingras Y 2022 Similarity network fusion for scholarly journals *J. Inform.* **16** 101226
Baccini F, Bianchini M and Geraci F 2022 Graph-based integration of histone modification profiles *Mathematics* **10** 1842
Banks D and Carley K 1994 Metric inference for social networks *J. Classif.* **11** 121–49
Bastian M, Heymann S and Jacomy M 2009 Gephi: an open source software for exploring and manipulating networks *Proc. Int. AAAI Conf. on Web and Social Media (San Jose, CA)* pp 361–2
Battiston F, Cencetti G, Iacopini I, Latora V, Lucas M, Patania A, Young J-G and Petri G 2020 Networks beyond pairwise interactions: structure and dynamics *Phys. Rep.* **874** 1–92
Berman A and Plemmons R J 1994 *Nonnegative Matrices in the Mathematical Sciences* (Philadelphia, PA: SIAM)
Berman A and Shaked-Monderer N 2003 *Completely Positive Matrices* (Singapore: World Scientific)
Bhatia R 2009 *Positive Definite Matrices* (Princeton, NJ: Princeton University Press)
Bhatia R and Congedo M 2019 Procrustes problems in Riemannian manifolds of positive definite matrices *Linear Algebr. Appl.* **563** 440–5
Bhatia R, Jain T and Lim Y 2019 On the Bures–Wasserstein distance between positive definite matrices *Expo. Math.* **37** 165–91
Bianconi G 2018 *Multilayer Networks: Structure and Function* (Oxford: Oxford University Press)
Blondel V D, Guillaume J-L, Lambiotte R and Lefebvre E 2008 Fast unfolding of communities in large networks *J. Stat. Mech.* P10008
De Domenico M 2022 *Multilayer Networks: Analysis and Visualization* (Cham: Springer)
Dickison M E, Magnani M and Rossi L 2016 *Multilayer Social Networks* (Cambridge: Cambridge University Press)
Fletcher P T and Joshi S 2007 Riemannian geometry for the statistical analysis of diffusion tensor data *Signal Process.* **87** 250–62
Huizing G-J, Deutschmann I M, Peyre G and Cantini L 2023 Paired single-cell multi-omics data integration with Mowgli *bioRxiv Preprint* pp 2023–02
Iannazzo B 2016 The geometric mean of two matrices from a computational viewpoint *Numer. Linear Algebr. Appl.* **23** 208–29
Jacomy M, Venturini T, Heymann S and Bastian M 2014 ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software *PLoS One* **9** e98679
Jeh G and Widom J 2002 SimRank: a measure of structural-context similarity *Proc. 8th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (Edmonton, Alberta)* pp 538–43
Johnson C R, Smith R L and Tsatsomeros M J 2020 *Matrix Positivity* (Cambridge: Cambridge University Press)
Kivelä M, Arenas A, Barthelemy M, Gleeson J P, Moreno Y and Porter M A 2014 Multilayer networks *J. Complex Netw.* **2** 203–71
Lambiotte R, Delvenne J-C and Barahona M 2008 Laplacian dynamics and multiscale modular structure in networks (arXiv:0812.1770)
Leydesdorff L 2008 On the normalization and visualization of author co-citation data: Salton's Cosine versus the Jaccard index *J. Am. Soc. Inf. Sci. Technol.* **59** 77–85
Liao X, Wu Y and Cao X 2019 Second-order CoSimRank for similarity measures in social networks *IEEE Int. Conf. on Communications (Shanghai, China)* (IEEE) pp 1–6
Lim Y and Pálfia M 2014a Weighted deterministic walks for the least squares mean on Hadamard spaces *Bull. London Math. Soc.* **46** 561–70
Lim Y and Pálfia M 2014b Weighted inductive means *Linear Algebr. Appl.* **453** 59–83
Lunagómez S, Olhede S C and Wolfe P J 2021 Modeling network populations via graph distances *J. Am. Stat. Assoc.* **116** 2023–40
Massart E M, Hendrickx J M and Absil P-A 2018 Matrix geometric means based on shuffled inductive sequences *Linear Algebr. Appl.* **542** 334–59
Newman M 2018 *Networks* (Oxford: Oxford University Press)
Omelka M and Hudecová S 2013 A comparison of the mantel test with a generalised distance covariance test *Environmetrics* **24** 449–60
Pavlopoulos G A, Kontou P I, Pavlopoulou A, Bouyioukos C, Markou E and Bagos P G 2018 Bipartite graphs in systems biology and medicine: a survey of methods and applications *GigaScience* **7** giy014
Robert P and Escoufier Y 1976 A unifying tool for linear multivariate statistical methods: the *RV*-coefficient *J. R. Stat. Soc.* C **25** 257–65
Shaked-Monderer N and Berman A 2021 *Copositive and Completely Positive Matrices* (Singapore: World Scientific)
Székely G J and Rizzo M L 2017 The energy of data *Annu. Rev. Stat. Appl.* **4** 447–79
Szekely G J and Rizzo M L 2023 *The Energy of Data and Distance Correlation* (Boca Raton, FL: CRC Press)
Székely G J, Rizzo M L and Bakirov N K 2007 Measuring and testing dependence by correlation of distances *Ann. Stat.* **35** 2769–94
Van Eck N J and Waltman L 2009 How to normalize cooccurrence data? An analysis of some well-known similarity measures *J. Am. Soc. Inf. Sci. Technol.* **60** 1635–51
Wang B, Mezlini A M, Demir F, Fiume M, Tu Z, Brudno M, Haibe-Kains B and Goldenberg A 2014 Similarity network fusion for aggregating data types on a genomic scale *Nat. Methods* **11** 333–7
Wasserman S and Faust K 1994 *Social Network Analysis: Methods and Applications* (Cambridge: Cambridge University Press)
Yger F, Berar M and Lotte F 2017 Riemannian approaches in brain-computer interfaces: a review *IEEE Trans. Neural Syst. Rehabil. Eng.* **25** 1753–62