Diffusion geometry of multiplex and interdependent systems

Giulia Bertagnolli [*]

CoMuNe Lab, Fondazione Bruno Kessler, Via Sommarive 18, 38123 Povo (TN), Italy
and University of Trento, Department of Mathematics, Via Sommarive, 14, 38123 Povo (TN), Italy

Manlio De Domenico [†]

CoMuNe Lab, Fondazione Bruno Kessler, Via Sommarive 18, 38123 Povo (TN), Italy

Complex networks are characterized by latent geometries induced by their topology or by the dynamics on them. In the latter case, different network-driven processes induce distinct geometric features that can be captured by adequate metrics. Random walks, a proxy for a broad spectrum of processes, from simple contagion to metastable synchronization and consensus, have been recently used, Domenico [Phys. Rev. Lett. **118**, 168301 (2017)] to define the class of diffusion geometries and pinpoint the functional mesoscale organization of complex networks from a genuine geometric perspective. Here we first extend this class to families of distinct random walk dynamics—including local and nonlocal information—on multilayer networks—a paradigm for biological, neural, social, transportation, and financial systems—overcoming limitations such as the presence of isolated nodes and disconnected components, typical of real-world networks. We then characterize the multilayer diffusion geometry of synthetic and empirical systems, highlighting the role played by different random search dynamics in shaping the geometric features of the corresponding diffusion manifolds.

I. INTRODUCTION

Complex networks, an abstract representation of the structural and functional backbone of complex systems, exhibit a wide spectrum of geometric features, from self-similarity to latent hidden metric spaces and topology [1–4], which have been successfully exploited to gain new insights about the structure and dynamics of social [5], neural [6], transportation [7,8], and communication systems [9], to mention a few emblematic examples (see [10] for a review). More recently, it has been shown that even the dynamics on complex networks can induce complex geometries which cannot be understood from inspecting only structural ones. Such geometries have been studied in the case of a broad class of spreading phenomena: communicability distance has been introduced to study information exchange based on walks [11]; an effective distance has been introduced for contagion processes [12]; a diffusion distance has been introduced to study collective phenomena such as synchronization, consensus and random searches [13]; and a temporal distance has been defined to study how perturbations spread in biological systems [14]. The common rationale is to model the propagation of information with network dynamics and investigate the induced distances between pairs of nodes.

Random walks are emblematic examples used for modeling diffusion and transport dynamics from lattices to disordered media and quantum systems across more than a century [15–21]. They provide both an intuitive—and often analytically treatable—mathematical framework and a rich physical model that can be used to map a wide spectrum of random processes and observed features, especially in the contest of complex networks, where the interplay between the underlying topology and the dynamics on it is responsible for the system's function [22]. The properties of the random walks reflect particular structural features of a system: for instance, they can be used to unravel the mesoscale organization of systems from a functional perspective [23–27]; localization effects can be observed around topological defects [28]; the average number of different sites visited in a time interval by a random walker, the so-called coverage, can be used to quantify the system resilience to random failures [29]; the relative importance of system's units can be quantified in terms of their ability to attract the overall flow, encoded by random walkers [30]; and the dynamical features of the information flowing through a network can be understood [31], and even enhanced [32], in terms of the spectral entropy defined by statistical features of random walks.

Based on these dynamics, one can also define similarity measures that reflect the ability of the units to exchange information [13,33–35]. In [13] diffusion distance was defined for single-layer networks and it has proven to be useful for characterizing the functional structure of complex networks, e.g., identifying functional clusters, or central nodes [36]. It also provides a real-valued distance function which, in applications where the global information about the underlying topology is missing or partial, is more informative than the purely topological length of shortest paths.

---

[*]Corresponding author: giulia.bertagnolli@unitn.it
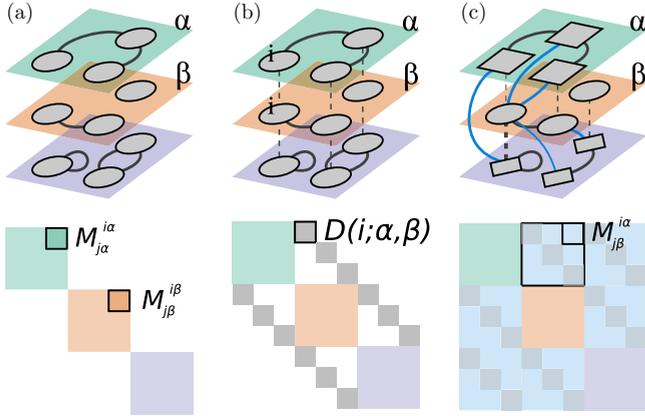[†]Corresponding author: mdedomenico@fbk.eu

FIG. 1. Different multilayer networks and their supra-adjacency matrix representation [37,40]: (a) an edge-colored multigraph, with layers corresponding to colors and no interlayer connections; (b) a multiplex network where the replicas in the different layers are interconnected sequentially, a type of intertwining or diagonal coupling; and (c) the most general interconnected case, where interlayer connections are not restricted to replicas (exogenous interactions). Latin letters denote nodes, and Greek letters are used for layers. $D(i; \alpha\beta)$ is the intensity of the intertwining between state nodes i, α and i, β. $M_{j\beta}^{i\alpha}$ describes the general interlayer connectivity.

However, in many real systems entities interact in multiple ways; think, for example, of locations connected through different transportation modes or to metabolites in a biological network linked by various types of chemical reactions. The evolution of the traditional graph-theoretical tools allows us nowadays to represent this multifaceted information by means of layers defining multilayer networks [37,38]. Figure 1 shows distinct representations of the multiplexity and the interdependency observed in empirical complex networks. From a mathematical perspective, taking into consideration the variety of connectivity patterns and coupling among layers and dynamics [39] requires multilinear algebra and tensors [37] to be efficiently represented and treated analytically.

In this study our contribution is twofold: on the one hand, we generalize multilayer RW dynamics for the analysis of more realistic empirical systems, e.g., allowing for the existence of isolated nodes and components; on the other hand, we extend the framework of diffusion geometry [13], that is, the study of the shape of networks embedded into space through diffusion dynamics, to the realm of multilayer systems, highlighting the dependence of the diffusion space on the interplay between structure and dynamics within and across layers.

The paper is organized as follows: in the next section we recall the basic facts on random walks and diffusion processes on single-layer networks and then generalize them to the multilayer framework. In Sec. III we extend the definition of diffusion distance to multilayer networks, presenting results obtained from the analysis of synthetic models. We conclude with applications to empirical social and transportation systems in Sec. IV, followed by discussion and conclusions.

II. DIFFUSION DYNAMICS IN COMPLEX NETWORKS

In a group of individuals, a piece of news spreads differently if everyone knows each other or if there are subgroups of individuals that do not communicate with the others: the structure of this social network influences the diffusion of information. Similarly, one observes a strong influence of the topology on the dynamics of a broad class of empirical systems. Because of this influence of the structure on the evolution of dynamical processes, the latter are often used as a proxy for probing and uncovering structural properties of complex networks. A simple strategy to explore a network is to start from a node, to choose at random an (outgoing) edge and to follow it toward the next node, i.e., to perform a random walk on the network with transition probabilities prescribed by the network connectivity. Random walks (RWs) represent a useful model for diffusion processes. Here we will focus on node-centric continuous-time random walks [22], continuous-time Markov chains on the vertex set of a network with transition rates depending on the network structure and specific navigation rules.

A. Random walks and diffusion distance on single-layer networks

Let us consider a weighted directed network $G = (V, E)$ without isolated nodes and with (possibly weighted) adjacency matrix $\mathbf{W} = (W_{ij}) \in \mathbb{R}^{N \times N}$, and two nodes $i, j \in V$. In a discrete classical RW the transition probability from i to j in one time step is given by $p_{ij} = \frac{W_{ij}}{s_i}$; the probability of the random walker being in node j after n steps is encoded in the jth component of the row vector $\mathbf{p}(n)$ and can be found by means of the equation $\mathbf{p}(n) = \mathbf{p}(n-1)\mathbf{D}^{-1}\mathbf{W}$, where $\mathbf{D}$ is the diagonal matrix of out-strengths $D_{ii} = s_i = \sum_{j=1}^{N} W_{ij}$. The continuous-time RW corresponding to this jump chain is described by the forward equation

$$\dot{\mathbf{p}}(t) = -\mathbf{p}(t)\tilde{\mathbf{L}} , \tag{1}$$

where $\tilde{\mathbf{L}} = \mathbf{I} - \mathbf{D}^{-1}\mathbf{W}$ is the RW normalized Laplacian and $-\tilde{\mathbf{L}}$ is the generator of the continuous-time Markov chain; see Appendix A for further details. The transition probabilities after time t are given by the solution of (1) with some initial condition $\mathbf{p}(0) = \mathbf{p}_0$, i.e., $p_{ij}(t) = (\mathbf{p}_0 e^{-t\tilde{\mathbf{L}}})_{ij}$. Rewriting (1) as a system with $\mathbf{P}(t)$ being a matrix and with the initial distribution given by the identity matrix, $\mathbf{P}(0) = \mathbf{I}$, we obtain the unique solution $\mathbf{P}(t) = e^{-t\tilde{\mathbf{L}}}$. Its ith row $(e^{-t\tilde{\mathbf{L}}})_{i\cdot} = \mathbf{p}(t|i)$ is the probability vector corresponding to $\mathbf{p}_0 = \mathbf{e}_i$ and $(e^{-t\tilde{\mathbf{L}}})_{ij}$ is the probability of being in j, after time t, starting in i with probability 1. The diffusion distance [13] is the L^2 norm of the difference between rows of the matrix $e^{-t\tilde{\mathbf{L}}}$, the difference between the two posterior distributions at time t obtained starting from i and j, respectively, i.e.,

$$D_t(i, j) = \|\mathbf{p}(t|i) - \mathbf{p}(t|j)\|_2. \tag{2}$$

At each fixed time $t \geqslant 0$, $D_t(i, j)$ defines a distance on the vertex set, which is bounded in $[0, \sqrt{2}]$ for all pairs of nodes, indeed, $\|\mathbf{p}(t|i)\|_2 \leqslant \|\mathbf{p}(t|i)\|_1 = 1$. Furthermore, two nodes $i \neq j$ are close w.r.t. D_t if there is a large probability that two random walkers starting in i and j respectively, can be found at the same node at time t [33], i.e., if the joint

probabilities $p_k(t \mid i)p_k(t \mid j)$ of two (independent) Markov chains being in the same state k at time t are large. See Appendix B for details. Furthermore, as the diffusion time t increases, and assuming that the walk is ergodic, $\mathbf{p}(t|i)$ will tend to the stationary distribution $\boldsymbol{\pi}$ and $D_t(i, j) \to 0$. The family of diffusion distances define a family of kinematic geometries based on a network-driven process [10]. It has also been shown [13] that t plays the role of a scale parameter and that the most persistent mesoscale structure of the network can be obtained averaging $(D_t)_{t>0}$ over time. The average diffusion distance will be indicated by $\bar{D}_t$.

Definition 1 (Diffusion space 1). The vertex set endowed with the diffusion distance, for a fixed value of t, is a metric space (V, D_t). Furthermore, a random walk dynamics maps V to a set of points in space. Indeed, given $i \in V$, $(e^{-t\tilde{\mathbf{L}}})_{i\cdot} = \mathbf{p}(t|i)$ is a vector in R^N and the diffusion distance at time t between a pair of nodes i, j is the Euclidean distance between the corresponding probability vectors. We call this embedding of the network nodes in $\mathbb{R}^N$ through the diffusion distance, the *diffusion space* or *diffusion manifold*.

On a given network, which is completely characterized by its adjacency matrix $\mathbf{W}$, we can define RWs with different flavors. This enables us, not only to model a wider range of physical spreading processes, but also to remove the rather restrictive assumption of connectedness, which is necessary for writing the transition matrix $\mathbf{D}^{-1}\mathbf{W}$ of the classical random walk. Consequently, also the family of diffusion distances can be generalized to different types of RWs, as well as to different types of networked systems. In the remainder of this section, we extend the diffusion distance to multilayer networks, additionally exploring the varying patterns induced by different RW dynamics [27,29]. Of course, there are many other types of RWs that are not mentioned here, e.g., multiplicative processes [41] or correlated random walks [42], which are often defined with a specific motivation. Nevertheless, our framework is very general and, given a transition matrix $\mathbf{T}$, can be effortlessly extended considering the continuous-time Markov chain having exponential holding times with rate one and $\mathbf{T}$ as jumping matrix; see Appendix A.

Before moving to the multilayer case, let us introduce the tensorial notation [37] for the monoplex G, which allows us to generalize the formalism to multilayer networks. The adjacency matrix $\mathbf{W}$ can be seen as a rank-2 tensor. In this paper, we will not use the covariant notation and the Einstein summation convention, so that W_j^i denotes the component (i, j) of the tensor. The master equation (1), and the corresponding initial condition, can be rewritten as

$$\dot{p}_j(t) = -\sum_{i=1}^{N} \tilde{L}_j^i p_i(t), \quad p_i(0) = q_i \qquad (3)$$

with $\tilde{L}_j^i = \delta_j^i - T_j^i$ indicating the component (i, j) of the random walk normalized Laplacian tensor, δ_j^i the Kronecker delta, and T_j^i the component (i, j) of the transition probability tensor.

B. Random walks on edge-colored multigraphs

A multilayer network is defined by a set of N nodes V interacting with each other in multiple ways, simultaneously. The different types of interaction can be encoded by colors and grouped into layers. Each node $i \in V$ exists at least in one layer and, if it exists in multiple layers $\alpha \in \{1, \dots, L\}$, we usually refer to (i, α) as a *replica* or *state node*, in contraposition to the physical node i. Interactions of the same color determine the intralayer connectivity, while the others are called interlayer connections. The simplest multilayer structure, depicted in Fig. 1(a), is the edge-colored multigraph, where there is no additional information (e.g., no order relation) on the set of layers $\{1, \dots, L\}$ or, equivalently, interlayer connectivity is not present.

Definition 2 (Adjacency matrices for edge-colored multigraphs). Given a edge-colored multigraph $G = (V, E, \{1, \dots, L\}, c)$, where V is the set of physical nodes, E is a multiset, $\{1, \dots, L\}$ is the set of colors, called layers from now on, and $c : E \to \{1, \dots, L\}$ is a surjective function assigning a color to each edge, we denote by $W_{j\alpha}^{i\alpha}$ the adjacency matrix in layer $\alpha = 1, \dots, L$.

There are, essentially, two ways to define RWs on these networks. One possibility is to allow the random walker to follow a sequence of edges with different colors: colored edges are then treated as multiple edges and the degrees of a node counts all the edges, regardless of their colors [43]. This choice is equivalent to aggregate the multilayer system to a single layer and perform a classical RW on it. This approach is not desirable in general, because one does not know *a priori* if, and to which extent, information is lost while aggregating the structure of layers will affect the results.

The second approach, used successfully in other applications [32], is to run independent dynamics on each layer and then integrate the transition matrices over the layers with the necessary normalization. Let us denote the transition probability from i to j in layer $\alpha = 1, \dots, L$ by $T_{j\alpha}^{i\alpha}$. Then the overall transition probability from i to j is obtained as follows:

$$\langle T_j^i \rangle = \sum_{\alpha=1}^{L} m_{i,\alpha} T_{j\alpha}^{i\alpha}, \qquad \sum_{\alpha=1}^{L} m_{i,\alpha} = 1 \quad \forall i \in V, \qquad (4)$$

where $m_{i,\alpha} \geqslant 0$ are weights enabling us to tune the importance of each layer, relative to node i in the final dynamics of the system. If $m_{i,\alpha} = \frac{1}{L}$ for all i and α, then $\langle T_j^i \rangle$ is simply the average transition matrix over all layers. However, if we choose $m_{i,\alpha} = \frac{1}{\mu_i} \mathbf{1}_{\{s_\alpha^{i\alpha} \neq 0\}}$, with $s_\alpha^{i\alpha}$ being the out-strength of vertex i in layer α and $\mu_i = \sum_{\alpha=1}^{L} \mathbf{1}_{\{s_\alpha^{i\alpha} \neq 0\}}$ the multiplicity of node i, i.e., the number of layers in which i is not isolated, then we can discard the effect of i being isolated in one or more layers.

This approach is more desirable than the first one, since it preserves more information related to diversity of connectivity patterns across layers. The main difference between the two approaches emerges when the edges are weighted, besides colored: *a priori* the scale (nominal, ordinal, ratio, etc.) of the weights could be different and an inattentive summation could lead to errors. In Eq. (4), instead, for each $i \in V$, we are taking a finite mixture of probability mass functions with weights $m_{i,\alpha} \geqslant 0$ such that $\sum_{\alpha=1}^{L} m_{i,\alpha} = 1$; nothing prevents one from

weighting layers differently, if additional information is available. In the following, we consider equal importance for all layers and choose $m_{i,\alpha} = \frac{1}{\mu_i} \mathbf{1}_{\{s_\alpha^{i\alpha} \neq 0\}}$.

Definition 3 (Transition matrix). Given a set V of N vertices a *transition matrix* $\mathbf{T}$ is an $N \times N$ matrix, whose entries T_j^i are non-negative reals representing the probability of moving from state (node, in the realm of graphs) i to j. $\mathbf{T}$ should also satisfy $\sum_{j=1}^N T_j^i = 1$. The ith row of the matrix $\mathbf{T}$ is then a probability vector, whose jth component is the probability of being in j given that the previous state was i. Given an edge-colored multigraph with state nodes $\{(i, \alpha) : i = 1, \dots, N$ and $\alpha = 1, \dots, L\}$ we define $T_{j\alpha}^{i\alpha}$ as the transition probability from (i, α) to (j, α). Transitions from (i, α) to (j, β) are not considered here, since there is no interlayer connectivity for edge-colored multigraphs and until now we just have a collection of L, $N \times N$ transition matrices.

We now take advantage of the more involved notation of Sec. II A to introduce multilayer interconnected networks.

C. Random walks on multilayer networks

In the same way a monoplex can be represented by a rank-2 tensor, a multilayer network is completely characterized by its rank-4 adjacency tensor [37]. Let us indicate by $M_{j\beta}^{i\alpha}$ the components of the (weighted) multilayer adjacency tensor. In this index notation $M_{j\beta}^{i\alpha}$ indicates the interaction between i and j in the same layer, while $M_{i\beta}^{i\alpha}$ denotes the strength of the intertwining of the replicas of i in two distinct layers. Depending on the interlayer connectivity we can have different types of multilayers, as shown in Fig. 1. When interlayer connections occur only between replicas of the same physical node, i.e., $M_{j\beta}^{i\alpha} = 0$ for $i \neq j$, the network is called a multiplex. The term $M_{i\beta}^{i\alpha}$ encodes the intertwining between two replicas; it is a scalar depending, in general, on i, α, β, and we indicate it by $D(i; \alpha, \beta)$, to emphasize the "diagonal coupling" [22,29,38] between replicas. For the ease of visualization, we flatten the adjacency tensor into the so-called supra-adjacency matrix [37,40,45], of dimension $NL \times NL$. Its diagonal blocks are the adjacency matrices of the monoplexes, while its off-diagonal blocks contain the information on the interlayer connectivity.

We henceforth indicate by $s_\alpha^{i\alpha} = \sum_{j=1}^N M_{j\alpha}^{i\alpha}$, the out-strength of node i in layer α and by $s^i = \sum_{\alpha=1}^L s_\alpha^{i\alpha} = \sum_\alpha \sum_j M_{j\alpha}^{i\alpha}$ its multilayer out-strength, discarding the interlayer edges; while, if interlayer edges are accounted for, $S^{i\alpha} = \sum_j \sum_\beta M_{j\beta}^{i\alpha}$ denotes the out-strength of node i in layer α. Observe that the interlayer strengths are then obtained as $S^{i\alpha} - s_\alpha^{i\alpha}$.

Generally, the presence of isolated nodes or components, as well as a heterogeneous or mixed distribution of interlayer connectivity is discarded to facilitate the analytical framework. Conversely, here we do not force any assumption on the structure of the multilayer system: if some units are not present in all layer—which is often the case for real data—we will add artificial replicas, which will appear as isolates and we will account for them adequately. We henceforth indicate by V the common vertex set and by $N = |V|$ the number of physical nodes. Although the isolated nodes do not modify

the connectivity of the connected component of the layer, they could be troublesome for the evaluation of transition probabilities. We have to distinguish three cases: (1) $S^{i\alpha} - s_\alpha^{i\alpha} = 0$, (2) $s_\alpha^{i\alpha} = 0$, and (3) $s^i = 0$. The last corresponds to the trivial case, where node i is isolated in every layer, so that i can be simply removed from the vertex set. We can assume, without loss of generality, that $s^i > 0$ for all $i \in V$. Cases (1) and (2), corresponding to no interlayer and no intralayer connections, respectively, constitute a problem only if $S^{i\alpha} = 0$ for some α. In this case one could see (i, α) as an absorbing state with the probability of remaining there equal to 1; another option is to teleport the random walker in (i, α) to any (j, β) with uniform probability $\frac{1}{NL}$. As for other approaches [23,27,30], we opted for the second option, since it decreases the occupation probability of the state node (i, α).

In the multilayer framework the probability transitions in one time step constitute the components $T_{j\beta}^{i\alpha}$ of a rank-4 tensor, and we can expand the RW equation to highlight different contributions of jumps and switches in the dynamics

$$
p_{j\beta}(t+1) = \underbrace{T_{j\beta}^{j\beta} p_{j\beta}(t)}_{\text{stay}} + \underbrace{\sum_{\substack{\alpha=1 \\ \alpha \neq \beta}}^{L} T_{j\beta}^{j\alpha} p_{j\alpha}(t)}_{\text{switch}}
$$

$$
+ \underbrace{\sum_{\substack{i=1 \\ i \neq j}}^{N} T_{j\beta}^{i\beta} p_{i\beta}(t)}_{\text{jump}} + \underbrace{\sum_{\substack{\alpha=1 \\ \alpha \neq \beta}}^{L} \sum_{\substack{i=1 \\ i \neq j}}^{N} T_{j\beta}^{i\alpha} p_{i\alpha}(t)}_{\text{switch and jump}} .
$$

Its continuous-time version is described by the forward equation

$$
\dot{p}_{j\beta}(t) = - \sum_{i,\alpha} \tilde{L}_{j\beta}^{i\alpha} p_{i\alpha}(t), \tag{5}
$$

where $\tilde{L}_{j\beta}^{i\alpha} = \delta_{j\beta}^{i\alpha} - T_{j\beta}^{i\alpha}$ and $p_{j\beta}(t)$ is the component of a rank-2 tensor indicating the probability of finding a random walker in node j of layer β at time t. When we want to highlight the initial condition, e.g., $p_{i\alpha}(0) = 1$ for the RW dynamics, we will use the conditional notation $p_{j\beta}[t|(i, \alpha)]$.

On a given network structure we can define different types of random walks, depending on the specific rules we want our random walker to explore the network. For instance, in PageRank [30] a *teleportation* or *jumping* parameter gives the possibility to the walker to reach also nodes that are not directly connected to the current node. The flavors of the random walks are given by their transition rules, which depend on the structure of the network, as a function of the adjacency tensor. The RW presented in the monoplex case is referred to as a classical random walk (CRW). Additionally to the CRW, we look here at a family of diffusion distances based on four other random walk types: multilayer PageRank (PRRW) generalizing [30,44], multilayer diffusive (DRW) generalizing the one defined in [29], maximal-entropy (MERW) generalizing [28,29], and physical random walk with relaxation (PrRW) [27], whose transition probabilities are shown in Table I.

(a) Classical (CRW), PageRank (PRRW), maximal-entropy (MERW) random walks. A classical random walker in a state node (i, α) of a multilayer network can move according to

TABLE I. Transition probabilities for different random walks: (CRW) classical, (PRRW) PageRank, (DRW) diffusive, (MERW) maximal-entropy, and (PrRW) physical with relaxation random walks. $s_{\max} = \max_{i,\alpha}\{S^{i\alpha}\}$; the jumping parameter of the PageRank RW is commonly indicated by α, so to avoid ambiguity, we indicate it by r; $\lambda_{\max}$ is the largest eigenvalue of the adjacency tensor and V is its corresponding eigentensor, satisfying $\sum_{i,\alpha} M^{i\alpha}_{j\beta} V_{i\alpha} = \lambda_{\max} V_{j\beta}$ (see [37] for details). Note that for CRW, DRW, and MERW, these transition rules generalize the one introduced in [29] for the analysis of multiplex networks. The PrRW is defined as in [27], while PRRW generalizes the walk introduced in [44].

	CRW [29]	PRRW [44]	DRW [29]	MERW [29]	PrRW [27]
$T^{j\beta}_{j\beta}$	$\frac{M^{j\beta}_{j\beta}}{S^{j\beta}}$	$r\frac{M^{j\beta}_{j\beta}}{S^{j\beta}} + (1-r)\frac{1}{NL}$	$\frac{s_{\max}+M^{j\beta}_{j\beta}-S^{j\beta}}{s_{\max}}$	$\frac{M^{j\beta}_{j\beta}}{\lambda_{\max}}$	$(1-r)\frac{M^{j\beta}_{j\beta}}{s^{j\beta}_\beta} + r\frac{M^{j\beta}_{j\beta}}{s^j}$
$T^{j\alpha}_{j\beta}$	$\frac{M^{j\alpha}_{j\beta}}{S^{j\alpha}}$	$r\frac{M^{j\alpha}_{j\beta}}{S^{j\alpha}} + (1-r)\frac{1}{NL}$	$\frac{M^{j\alpha}_{j\beta}}{s_{\max}}$	$\frac{M^{j\alpha}_{j\beta}}{\lambda_{\max}}\frac{V_{j\beta}}{V_{j\alpha}}$	$r\frac{M^{j\beta}_{j\beta}}{s^j}$
$T^{i\beta}_{j\beta}$	$\frac{M^{i\beta}_{j\beta}}{S^{i\beta}}$	$r\frac{M^{i\beta}_{j\beta}}{S^{i\beta}} + (1-r)\frac{1}{NL}$	$\frac{M^{i\beta}_{j\beta}}{s_{\max}}$	$\frac{M^{i\beta}_{j\beta}}{\lambda_{\max}}\frac{V_{j\beta}}{V_{i\beta}}$	$(1-r)\frac{M^{i\beta}_{j\beta}}{s^{i\beta}_\beta} + r\frac{M^{i\beta}_{j\beta}}{s^i}$
$T^{i\alpha}_{j\beta}$	$\frac{M^{i\alpha}_{j\beta}}{S^{i\alpha}}$	$r\frac{M^{i\alpha}_{j\beta}}{S^{i\alpha}} + (1-r)\frac{1}{NL}$	$\frac{M^{i\alpha}_{j\beta}}{s_{\max}}$	$\frac{M^{i\alpha}_{j\beta}}{\lambda_{\max}}\frac{V_{j\beta}}{V_{i\alpha}}$	$r\frac{M^{i\beta}_{j\beta}}{s^i}$

intra- or interlayer connections, with uniform probability. The normalizing factor for the transition probabilities is then the total strength $S^{i\alpha}$. The PRRW differs from the CRW in that the walker moves with probability r according to its edges (of both types) and with probability $1-r$ it may jump (or teleport itself) to any state node in the network. In MERW, as can be seen in Table I, the transition probabilities are governed by the largest eigenvalue, and corresponding eigenvector, of the adjacency tensor. Here we assume the networks to be undirected. In the discrete-time framework, it has been shown [28] that all the trajectories of length l between two given nodes have the same probability and that the MERW has maximal entropy production rate among all walks on the network.

(b) Diffusive random walk (DRW). The jumping process described by the transition probabilities in DRW Table I is characterized by a nonzero probability of remaining in a given state node (i,β), $1 - \frac{S^{i\beta}}{s_{\max}}$, and the probability of following out-going links with a probability normalized over the maximum strength over all nodes, $s_{\max}$.

(c) Physical random walk with relaxation (PrRW). The physical RW has been defined in [29] to describe those dynamics where the state nodes have a "common memory," so that the information diffuses instantaneously across replicas. Think, for instance, of the system of virtual interactions among individuals, who may have a profile (an alter-ego) in different social networks. A person can then exchange information in a particular social network using the (intralayer) connections of its alter-ego in that social system, but she or he has always a complete knowledge of the information across the layers. In this case, interlayer connections between replicas of different physical nodes have no physical meaning and, consequently, are ignored. The PrRW [27] is a variant on the physical random walk, where the assumption on the complete knowledge of intertwining between layers is dropped. It can be seen from Table I that its transition probabilities contain a trade-off between intra- and interlinks, which are followed with probability $1-r$ and r, respectively. If not differently stated, we consider here $r = 0.5$.

In the following section we will define a multilayer distance which reflects how nodes exchange information in and between the layers.

III. DIFFUSION DISTANCE IN MULTILAYER SYSTEMS

In the multilayer framework the probability of finding a random walker at a given node and layer is encoded in a time-dependent tensor, whose component $p_{j\beta}(t)$, corresponds to the probability of finding the random walker in the state (j,β), for a given initial distribution. Similarly to (2), we can then define the diffusion distance between state nodes (i,α) and (j,β) as

$$D^2_t((i,\alpha),(j,\beta)) = \sum_{k,\gamma}\{p_{k\gamma}[t|(i,\alpha)] - p_{k\gamma}[t|(j,\beta)]\}^2. \quad (6)$$

As for the monoplex case, we can write the transition probabilities in terms of the exponential of the supra-Laplacian matrix, so that Eq. (6) can be written as

$$D^2_t((i,\alpha),(j,\beta)) = \sum_{k,\gamma}\left[(e^{-t\tilde{L}})^{i\alpha}_{k\gamma} - (e^{-t\tilde{L}})^{j\beta}_{k\gamma}\right]^2 \quad (7)$$

and can be characterized by the spectrum of the Laplacian tensor [37,40]. At t fixed, $D^{i\alpha}_{j\beta}(t) = D_t((i,\alpha),(j,\beta))$ is the component of a metric tensor, which can be flattened into a $NL \times NL$ matrix indicated by $\mathbf{D}_t$ and called the supradistance matrix, using the same nomenclature of "supramatrices" introduced in [38]. Let us now summarize the just introduced concepts:

Definition 4 (Diffusion distance, tensor, supradistance matrix, average diffusion distance, and normalized supradistance matrix.). Given a multilayer network and a RW dynamics on it, whose transition probabilities at time t are given by $p_{j\beta}[t|(i,\alpha)] = (e^{-t\tilde{L}})^{i\alpha}_{j\beta}$, we define the diffusion distance between state nodes (i,α) and (j,β) at time t, $D_t((i,\alpha),(j,\beta))$, as in Eq. (7). The terms of the resulting diffusion distance tensor are denoted by $D^{i\alpha}_{j\beta}(t) = D_t((i,\alpha),(j,\beta))$. We denote by $\mathbf{D}_t$ the diffusion supradistance matrix, resulting from the flattening of the rank-4 tensor. The average diffusion distance [13] is defined as

$$\bar{D}_t = \frac{1}{t_{\max}}\sum_{\tau=1}^{t_{\max}} D_\tau,$$

where $t_{\max}$ is a temporal cutoff. The corresponding average diffusion supradistance matrix is indicated by $\bar{\mathbf{D}}_t$. Finally, we
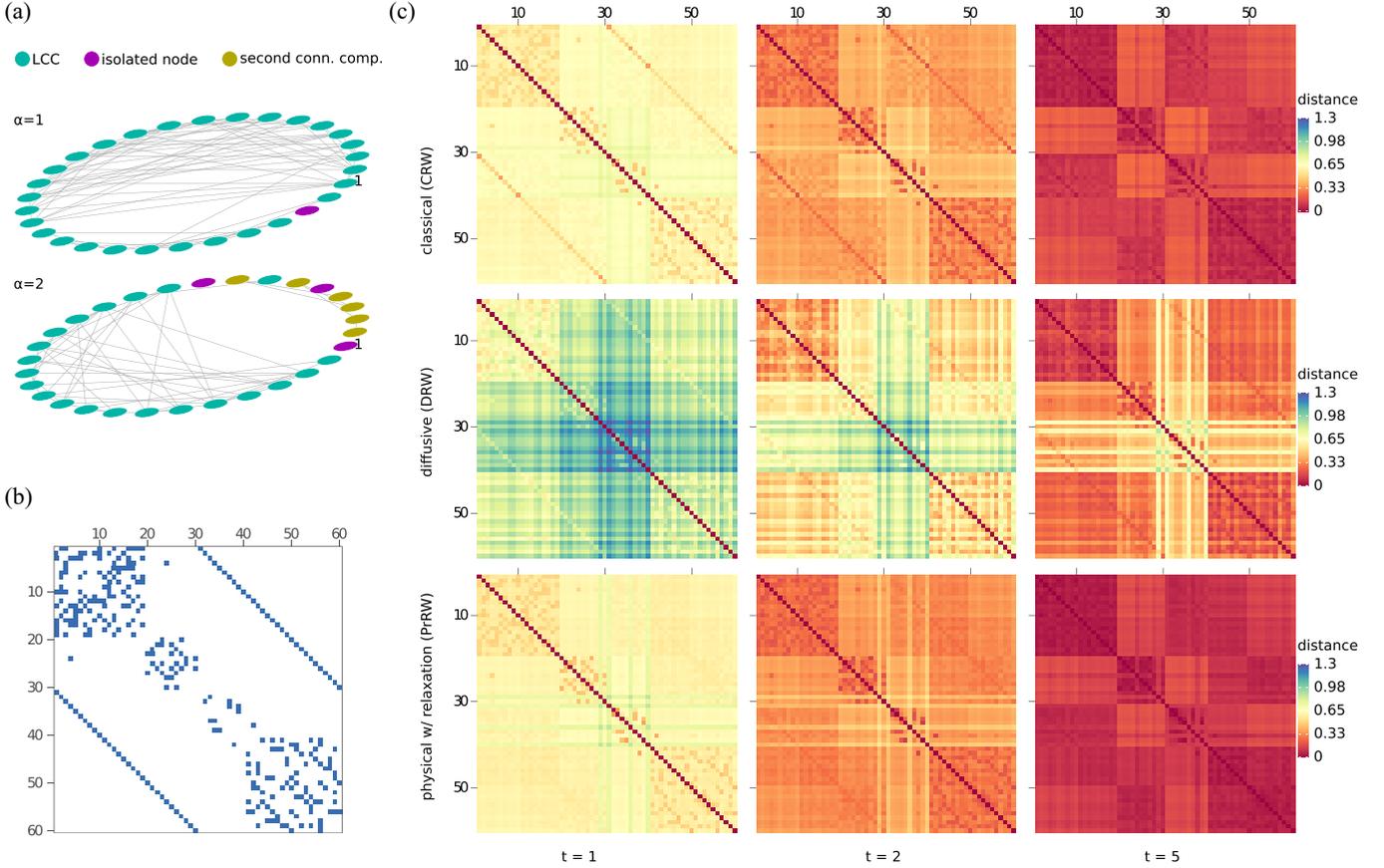
FIG. 2. A synthetic two-layer network with unitary diagonal coupling $D(i; \alpha, \beta) = 1$, for all i and $\alpha \neq \beta$, and its functional characterization. (a) Isolated nodes, the largest connected component (LCC), and other different components have been highlighted in the two layers. (b) The supra-adjacency matrix of the multiplex. (c) The $NL \times NL$ diffusion supradistance matrices for three different RW dynamics and $t = 1, 2, 5$. The PrRW is evaluated for $r = 0.5$.

introduce the normalized diffusion supradistance matrix [13]

$$\tilde{\mathbf{D}}_t = \frac{\mathbf{D}_t}{\max \mathbf{D}_t}.$$

Figure 2 shows the supradistance matrix $\mathbf{D}_t$ for the three more diverse RW dynamics on a synthetic multilayer network. In each layer, we have a network with $N = 30$ nodes generated from a stochastic block model with two blocks. We chose the probabilities in order to have a diverse topology: dense groups, disconnected components, and isolated nodes. The coupling between the layers is $D(i; 1, 2) = D(i; 2, 1) = 1$, as shown in the supra-adjacency matrix of Fig. 2(b). The columns of panel (c) represent different diffusion times t. Recall that the diffusion time plays the role of a scale parameter [13] and that the continuous-time Markov chain has exponentially distributed holding times with rate $\lambda = 1$, i.e., the expected time occurring among each step of the RW is 1. $D_{t=1}$ is then a function of the microscale structure of the multiplex and here the isolated nodes are clearly visible. Remarkably, the distances w.r.t. the DRW span a larger interval than the others and differentiate very effectively the subgroups in each layer. This is a consequence of the diffusive dynamics, where the jumping probabilities do not depend on the vertex, but on the strength of hubs. It is also worth noticing that those nodes, which are disconnected from the largest connected

component of the second layer, are generally nearer to the nodes in the first layer than to the nodes in their same layer. To unveil the persistent mesoscale structures of the network we averaged the diffusion supradistance matrices for up to $t_{\max} = N$ [13] and run a hierarchical clustering on the average diffusion supradistance matrices, summarizing the results in Fig. 3. As expected, the presence of interlinks moves loosely connected nodes closer to their replica in the other layer and the clustering does not separate the two layers. In this and the following plots, the state nodes in the supradistance matrices are reordered according to the dendrogram returned by the complete linkage method for hierarchical clustering, (R) HCLUST [46]. At the beginning, the two nearest nodes are grouped together, then at each step, the distance between the clusters is defined as the maximum distance between their nodes and the two nearest clusters are aggregated, until all the nodes are in the same group. In [13] it has been shown that the dichotomy between fast-shrinking (intracommunity) and slow-shrinking (intercommunities) distances probes the mesoscale structure at different resolutions and that the most persistent and representative mesoscale structure is the one maximizing the average diffusion distance between clusters. Here we use the dendrograms just as a guide for the eye and do not delve into the multilayer community detection [27,47] or the analysis of the hierarchies [48].
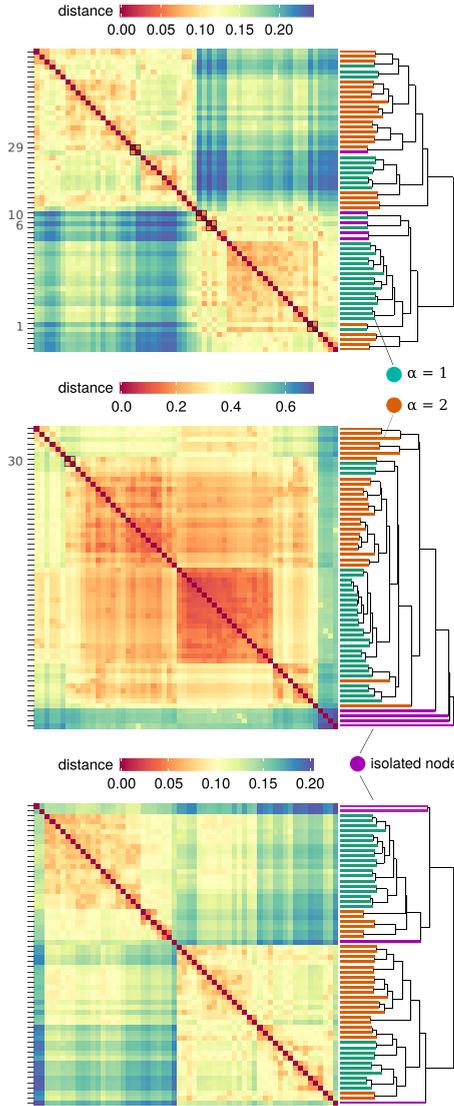
FIG. 3. Average diffusion distance matrices of the two-layer synthetic network shown in Fig. 2(a), w.r.t. three RW dynamics: (top) classical, (middle) diffusive, and (bottom) physical with relaxation $r = 0.5$. The state nodes (i, α) are colored in the corresponding dendrogram according to the layer they belong to, except for isolated nodes, highlighted in purple. Small squares with a black border indicate those state nodes grouped together, e.g., $(29, 1), (29, 2)$ in the top panel.

Finally, we provide a grounded way to summarize the supradistance matrix into an $N \times N$ matrix collecting the diffusion distances among the physical nodes, regardless of the layers. The ij elements of the diagonal blocks of the supramatrix, $\{D_t((i, \alpha), (j, \alpha)), \alpha \in 1, \ldots, L\}$, can be seen, using the jargon of electrical circuits, as resistances in parallel between the physical nodes i, j. Their equivalent resistance can then be found through the parallel sum of the resistances between the replicas, yielding to the equivalent diffusion distance:

$$D_t^{\text{eq}}(i, j) = \left[\sum_{\alpha=1}^{L} \frac{1}{D_t((i, \alpha), (j, \alpha))} \right]^{-1}. \tag{8}$$

Note the equivalent diffusion distance (briefly, *equivalent distance*, if not ambiguous) is as an aggregation of diffusion distances across layers, and it is not related to the concept of resistance distance [49].

The corresponding equivalent distance matrix, denoted by $\mathbf{D}_t^{\text{eq}}$, is quite different from the one obtained evaluating the diffusion distance on the aggregated network, as shown in Fig. 4. Here the distance matrices have been rescaled in $[0, 1]$ by division through their respective maximum values, reported in the figure caption. Looking back at Fig. 2(a), we can see that the nodes from 1 to 19 are densely connected in layer 1, while node 8 and nodes from 11 to 30 form the largest connected component of the second layer. In both distance matrices there is a clear block corresponding to the last 10 nodes. However, only the equivalent diffusion distance matrix captures the particular position of the first 10 nodes (removing 8): in the multiplex they are distant from the last 10, because they belong to different communities in layer 1 and to disconnected components in layer 2. We also compare the aggregated network distance matrix Fig. 4(b) to the one of the edge-colored multigraph obtained removing the interlayer links from the multilayer network, shown in Fig. 4(c). The two matrices are very similar, with only minor permutations inside the clusters.

As in the single-layer case, through RW dynamics we can map state nodes to points in space: indeed $p_{j\beta}(t) = \tilde{p}_l(t)$, with $l = j + \beta$ can also be seen as the component of a supravector in R^{NL}; see Appendix B. We then generalize Definition 1 as follows:

Definition 5 (Diffusion space 2). The set of state nodes $\{(i, \alpha) : i = 1, \ldots, N, \alpha = 1, \ldots, L\}$ endowed with the diffusion distance at a fixed value of t, is a metric space. A random walk dynamics maps each state node (i, α) to a point in the space $\mathbb{R}^{NL}$. Furthermore, as the Euclidean norm of the probability supravectors, with components $p_{j\beta}[t|(i, \alpha)]$, is bounded, the state nodes are embedded into a bounded subspace of $\mathbb{R}^{NL}$, which is here referred to as the *diffusion space* or *diffusion manifold* or *diffusion embedding*.

Multilayer diffusion manifolds

The use of different random walk dynamics to explore a system has an impact on the distances between its units and, consequently, on how the units are distributed in the induced diffusion spaces. Similarly, the diffusion time shapes the pairwise distances, highlighting local features of complex network geometry on short timescales and its more persistent structures for large diffusion times. In the multilayer setting there is an additional level of complexity given by the interlayer connections and by the layer-layer correlations. To gain further insights, we generate three distinct classes of synthetic multilayer networks, with system size $N = 200$, and analyze them through the lens of diffusion geometry.

The first class consists of Barabasi-Albert scale-free networks [51] on each layer: we consider a linear preferential attachment with four edges added by each new node during the growth process, while setting at 10% the edge overlapping across layers—defined in terms of the fraction of links which are present in both layers among the same pairs of nodes [52].
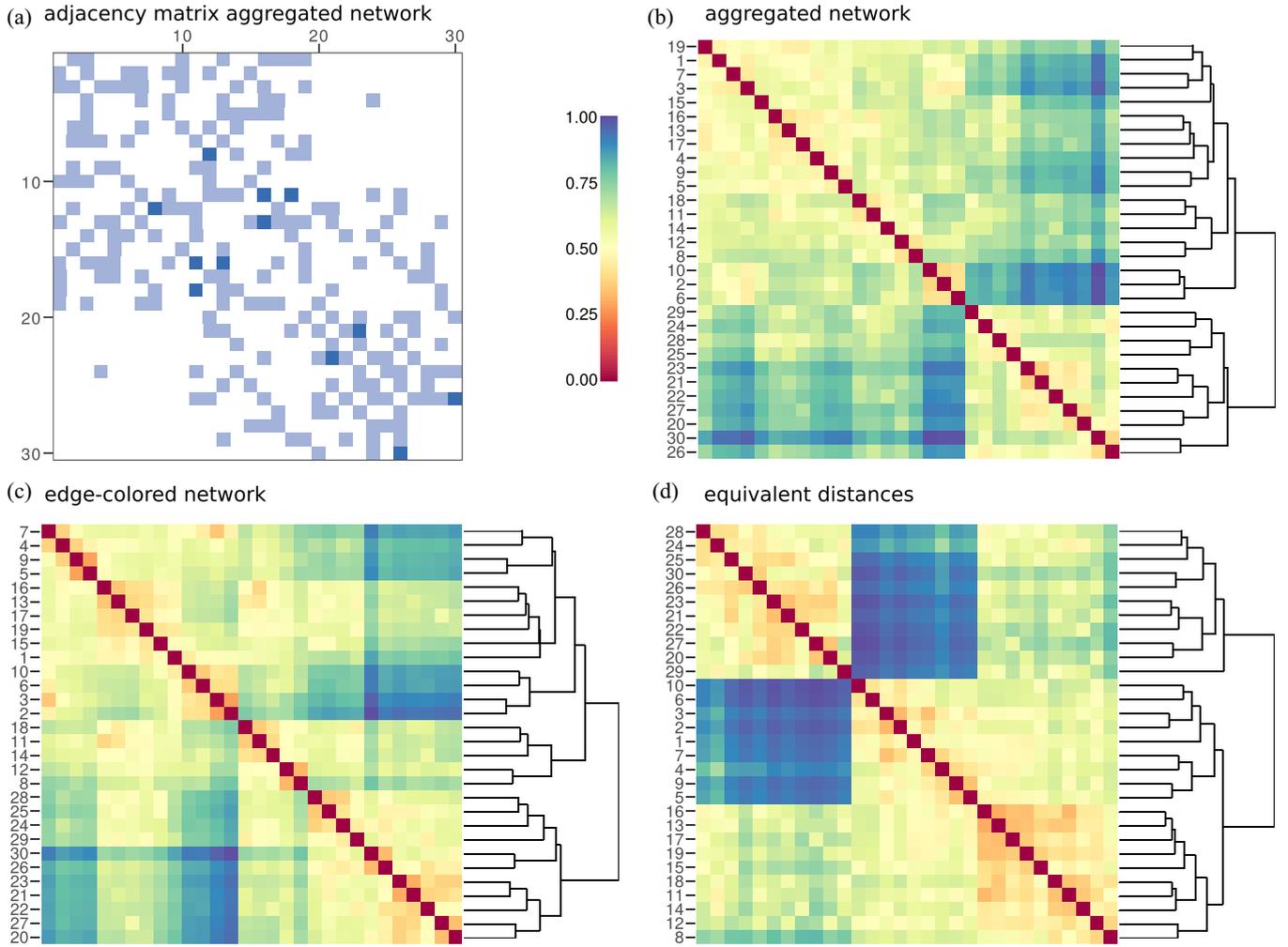
FIG. 4. Comparing different types of $(N \times N)$ average diffusion distance matrices between physical nodes, regardless of layers. (a) The weighted adjacency matrix of the aggregated network. The diffusion distance matrix (w.r.t. the CRW and averaged over time) (b) of the aggregated network corresponding to the synthetic multiplex of Figs. 2(a) and 2(c) of the edge-colored network. (d) The equivalent diffusion distance matrix obtained with the reduction of Eq. (8). The distances are rescaled to [0, 1], normalizing by the corresponding maximum over all node pairs [aggregated: $\max(\bar{\mathbf{D}}_t) \approx 0.13$, edge-colored: $\max(\bar{\mathbf{D}}_t) \approx 0.16$, equivalent distances: $\max(\bar{\mathbf{D}}_t^{\mathrm{eq}}) \approx 0.11$].

The second class consists of Watts-Strogatz small-world network [53] on each layer, obtained by rewiring lattices with probability 0.2, where edge overlapping is tuned similarly to the first class.

The third class consists of layers with strong mesoscale structure organized in four groups, like in a Girvan-Newman model [54] on each layer: the probability that two nodes within the same group are connected is 1, whereas cross-group connections are much sparser and present with probability 0.05. Group overlapping [52]—defined in terms of the fraction of nodes planted in the same group on both layers—is fixed at 1%.

The role of layer-layer correlations and their interplay with the distinct network topologies considered above is summarized in Fig. 5. As expected, there are relevant differences due to the type of random search dynamics and to the topological features of the underlying topologies. For instance, the diffusive walk for the Barabasi-Albert system leads to a high level of mixed pathways across layers. In the resulting diffusion

embedding the nodes are not grouped together according to their layer assignment, as can be seen from the mix of green and purple in the corresponding dendrogram of Fig. 5. For the same walk, in the case of the Watts-Strogatz system the result is the opposite: nodes aggregate into highly homogeneous functional clusters w.r.t. the nodes layer assignment. As expected, a high amount of *geometric mixing* (i.e., the mixing in the embedding space of state nodes that belong to different layers or communities in the network) is also observed when the network functionality is studied through the PrRW. This dynamics on the Girvan-Newman model makes intra- and interlayer distances more homogeneous than the other RWs and the cross-layer communities appear only for very large values of partition overlap; see Fig. 11 in Appendix D. Overall, it is not guaranteed that diffusion pathways across layers favor the geometric mixing in the diffusion manifold: the result depends on the type of dynamics and on layer-layer correlations.

Delving deeper into this interplay between structure and dynamics, we have considered a second battery of synthetic
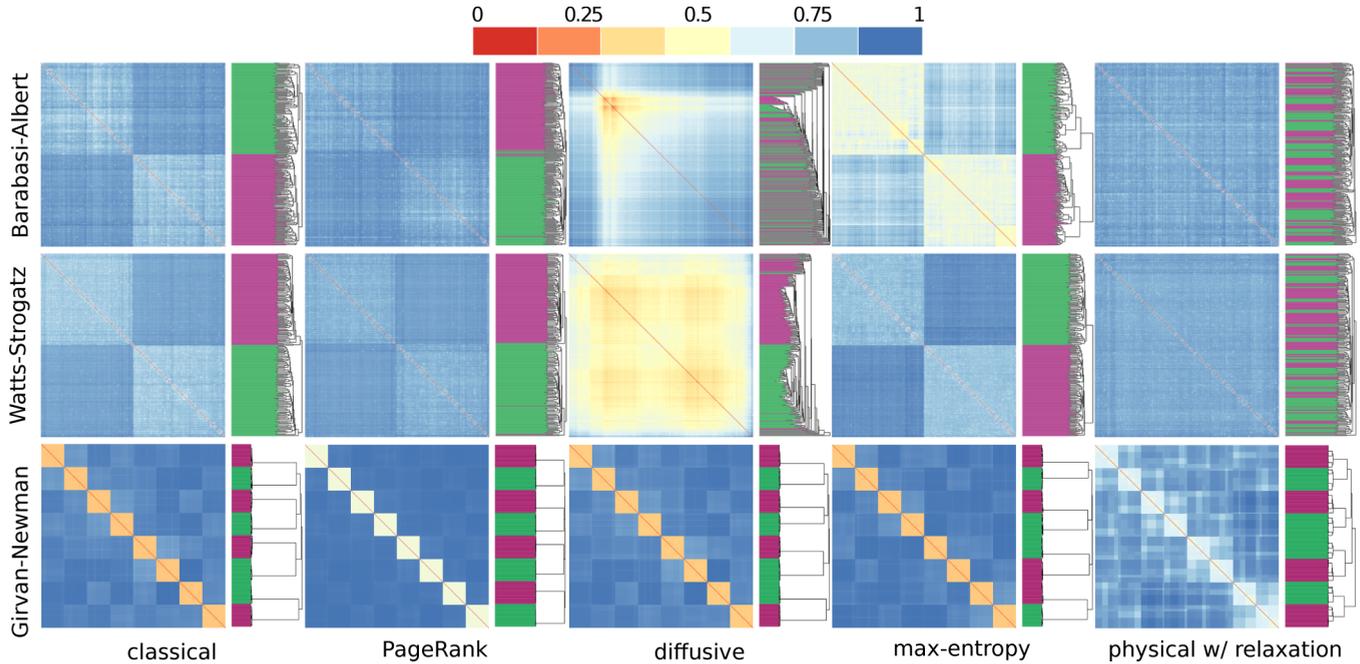
FIG. 5. Average diffusion distance $\bar{D}_t$ on two-layer multiplexes with different topologies, for fixed values of global average edge overlap (Barabasi-Albert and Watts-Strogatz) and partition overlap (Girvan-Newman) between layers. Dendrograms on the right-hand side of each distance matrix represent the corresponding hierarchical clustering to highlight the mesoscale organization of the system, with color encoding the planted node assignment in each layer. Distances have been rescaled $\frac{\bar{D}_t}{\max \bar{D}_t}$. The Barabasi-Albert model is characterized by the presence of hubs, which are clearly recognizable in the supradistance matrix w.r.t. the DRW, despite the small overlap between layers. The Girvan-Newman two-layer multiplex has a mesoscale organized in strong communities, partly overlapping across layers. Note that at variance with edge overlapping, here partition overlapping is defined in terms of nodes belonging to the same group without requiring those nodes to be connected by an overlapping edge.

models, where we increasingly add interlayer connectivity between layers.

The absence of information pathways across layers, happening, for instance, when two layers are not coupled together, leads naturally to disjoint diffusion manifolds, i.e., disjoint clouds of points in the diffusion space, each one corresponding to the distinct layers. When the two layers are interconnected together, a trivial result is that the strength of interlayer connectivity facilitates the flow of information across layers. However, the above process hides an interesting phenomenon, which is unveiled in Figs. 6 and 7. To better characterize it, we calculate the Frobenius norm, which is defined as follows, for a generic matrix **A**,

$$\|\mathbf{A}\|_{\mathrm{F}} = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |A_{ij}|^2} = \sqrt{\mathrm{trace}\,(\mathbf{A}^T \mathbf{A})},$$

to quantify the overall intensity of an average diffusion distance matrix. The Frobenius norm is computed when the two layers are not coupled, on the union of the two distance matrices, i.e., $\sqrt{\|\mathbf{A}\|_{\mathrm{F}}^2 + \|\mathbf{B}\|_{\mathrm{F}}^2}$, and then on the supradistance matrices for increasing fraction of interlayer connectivity: first, state nodes corresponding to the same physical node are interconnected with each other to create an interconnected multiplex; after this regime is reached, the cross-links between state nodes corresponding to all other physical nodes are created, until the total of N^2 connections is generated.

Remarkably, when one interlink is added between the layers, the Frobenius norm increases: this is due to the fact that the new link coupling the two layers creates a bottleneck for the information exchange across layers, even for large values of t. Once more interlinks are added, the Frobenius norm decreases, tending to a plateau when the fraction of interlayer links approaches the 100%. It is worth noticing how the structures of the single layers and of the whole fully interconnected multilayer coexist in the different diffusion spaces w.r.t. the diffusion time t. In case of Barabasi-Albert networks in each monoplex, Fig. 6, their typical heterogeneity (presence of hubs) characterizes also the diffusion manifolds. As a matter of fact, hubs have smaller diffusion distances to any other node. On the other hand, for small values of t we can see in Fig. 7 that the community structure is concealed by the presence of the interlayer links. This is totally expected, since diffusion time plays the role of a scale parameter, and locally, nodes in the fully interconnected multilayer with communities, Fig. 7(d), are very similar.

Our results highlight that the existence of topological correlations across layers induce changes in how information is exchanged between state nodes. Such changes alter diffusion distances and might lead to two different regimes: (1) flow keeps segregated within layers and the diffusion manifold corresponding to the multilayer consists of two well-separated submanifolds representing each layer separately and (2) flow is integrated and the diffusion pathways across layers mix up those submanifolds.
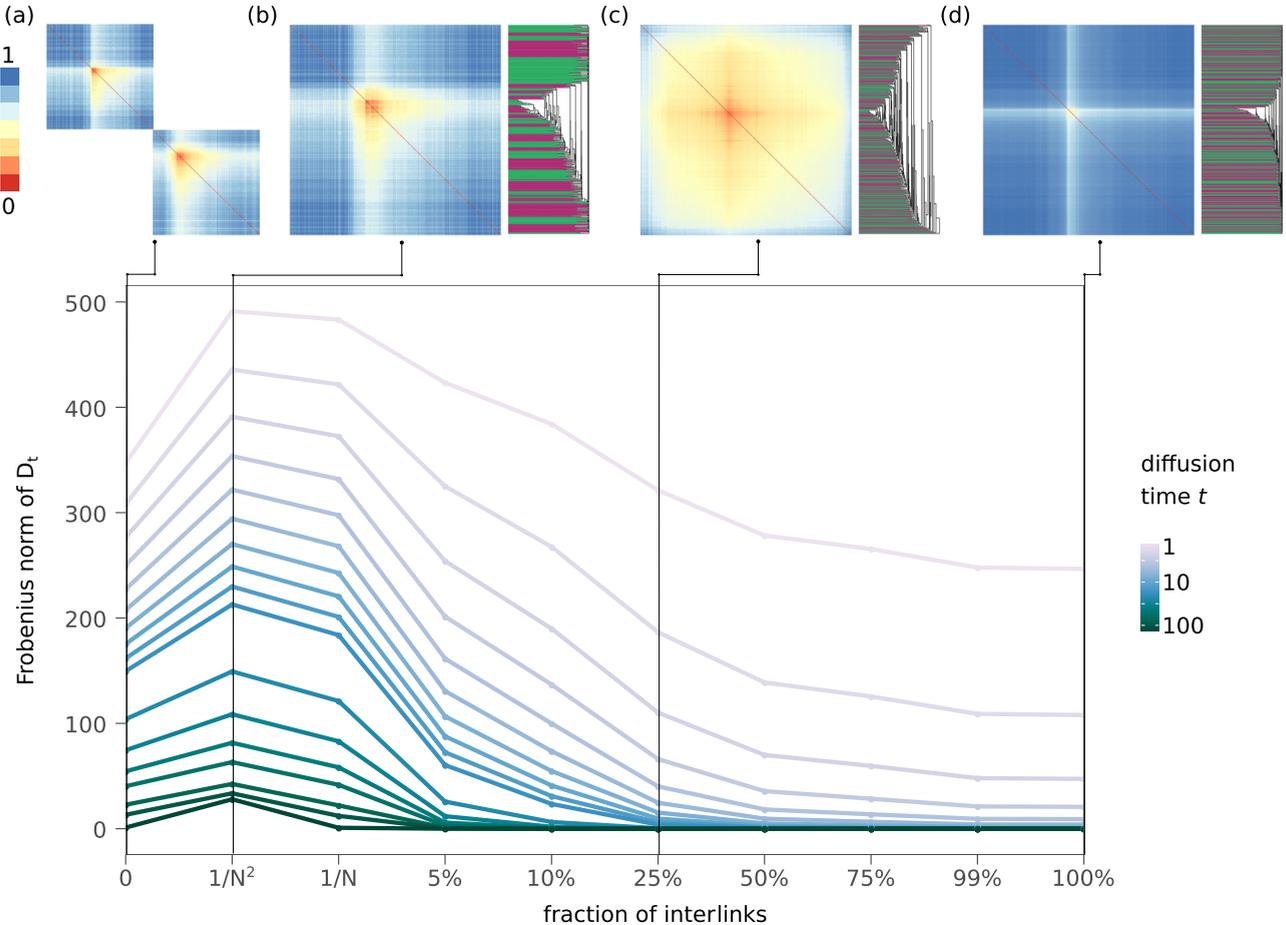
FIG. 6. Frobenius norm of the supradistance matrices of a synthetic two-layer network, w.r.t. a DRW. Each layer is generated from a Barabasi-Albert model as in Fig. 5. As the fraction of interlayer links grows we move from two disconnected multiplexes to a fully interconnected multilayer whit all N^2 connections across layers. The heatmaps of $D_{t=5}$ are four representatives of the different regimes: (a) uncoupled layers; (b) a single-interlink between the replicas of a random node coupling the two layers, i.e., $[\frac{1}{N^2}, \frac{1}{N}]$ partially interconnected multiplex, and $[1/N]$ fully interconnected multiplex (all state nodes corresponding to the same physical node are interconnected); (c–d) $[\frac{1}{N}, 1]$ multilayer regime consisting of an interconnected multiplex with the addition of cross-links between state nodes of distinct physical nodes.

IV. APPLICATIONS TO EMPIRICAL MULTILAYER SYSTEMS

We use the introduced family of metrics to study two real systems with multiple types of interactions: the multimodal transportation network of London [29] and the multilayer No-ordin Top Terrorists Network [55]. The first system consists of three layers corresponding to the Tube, overground, and DLR (Docklands Light Railway), arranged in a multiplex with couplings $D(i; \alpha, \beta) = 1$ for $\alpha \neq \beta \in \{1, 2, 3\}$. Nodes represent stations ($N = 369$ in total) and connections between them are weighted and undirected. The network of interactions among 78 Indonesian terrorists ($N = 79$ in the data set, but actor 58 is usually removed since it is disconnected in all layers) is a four-layers multiplex, representing their pairwise trust (T), operational (O), communication (C) ties, and business (B) relations [56].

Figure 8 shows the average diffusion manifolds, i.e. the diffusion manifold corresponding to $\bar{D}_t$, projected in $\mathbb{R}^3$ through multidimensional scaling, induced by different RW dynamics. As observed in [29], the best exploration strategy, i.e., the RW to adopt to cover efficiently the network, depends

on the topology of the multilayer. This is reflected in the maps of Fig. 8, even though they are low-dimensional approximations of the true diffusion manifolds. As a matter of fact, for the London transportation network, the manifolds induced by the classical, PageRank, and diffusive RWs appear qualitatively very similar with each other, and considerably different from those induced by MERW and PrRW. Instead, the supradistance matrices and manifolds obtained for the terrorists network appear similar in that all have a group of nodes with small pairwise distances, and another group of nodes which are distant from each other. To quantify more adequately how diverse, or similar, the manifolds are, we compare their supradistance matrices by means of Mantel's test [57,58], where the null hypothesis is that the pairwise distances in one matrix are not monotonically related to the corresponding distances in the second matrix, and show the results of our test in Figs. 9 and 10.

V. DISCUSSION AND CONCLUSIONS

We have considered different families of random walk dynamics, adequately extended to the realm of multilayer
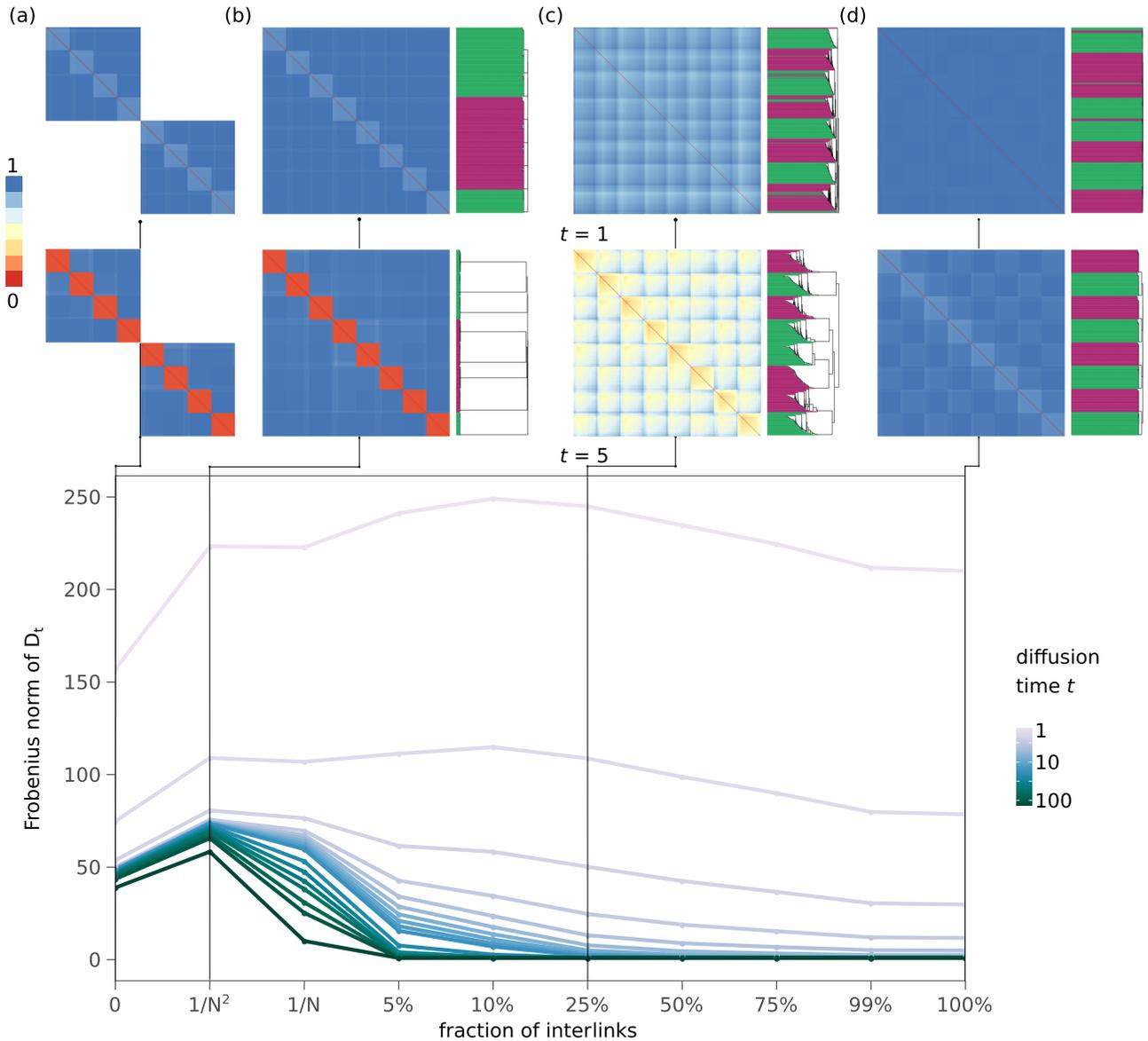
FIG. 7. As in Fig. 6 for networks with a mesoscale organized in four strong communities in each layer and partition overlap of 1%. The heatmaps on the top correspond to $D_{t=1}$, while the others are $D_{t=5}$.

networks, to introduce the multilayer diffusion geometry. Its classical counterpart, single-layer diffusion geometry, intimately relates metastable synchronization, consensus and random search dynamics, providing a framework for identifying functional clusters in complex networks.

While the framework and its validity remain the same, its natural generalization to multilayer networks—i.e., systems consisting of multiple types of relationships among their units—was missing. Here we fill this gap and provide evidence that multilayer diffusion distance and induced manifolds encode information due to the both the structure of the multilayer and the type of dynamics on it. Of course, here we focus only on averaging operators (Laplacians) and diffusion processes for studying the function of networks, while other dynamics may display different functional behaviors [59]. In the same spirit, different representations of complex systems accounting for other types of interactions, e.g., hypergraphs,

may provide diverse insights on the functionality of complex systems [60].

From the analysis of synthetic networks with overlapping edges or groups across layers, we have found that the interplay between dynamics and topology cannot be easily decoupled: e.g., the classical random walk reveals a manifold in which state nodes of the different layers have larger distances than the intralayer distances, but this does not remain true in the presence of strong communities, which are not necessarily overlapping. Also the behavior of the MERW is not trivial: the top-level hierarchical structure unveiled in the Barabasi-Albert is compatible with that of the Watts-Strogatz network, despite the high heterogeneity of the first, and this could be surprising since MERW is influenced by irregularities in nodes degree. In another scenario, where two layers are originally uncoupled and do not exchange information, we add interlayer connectivity to better understand how the
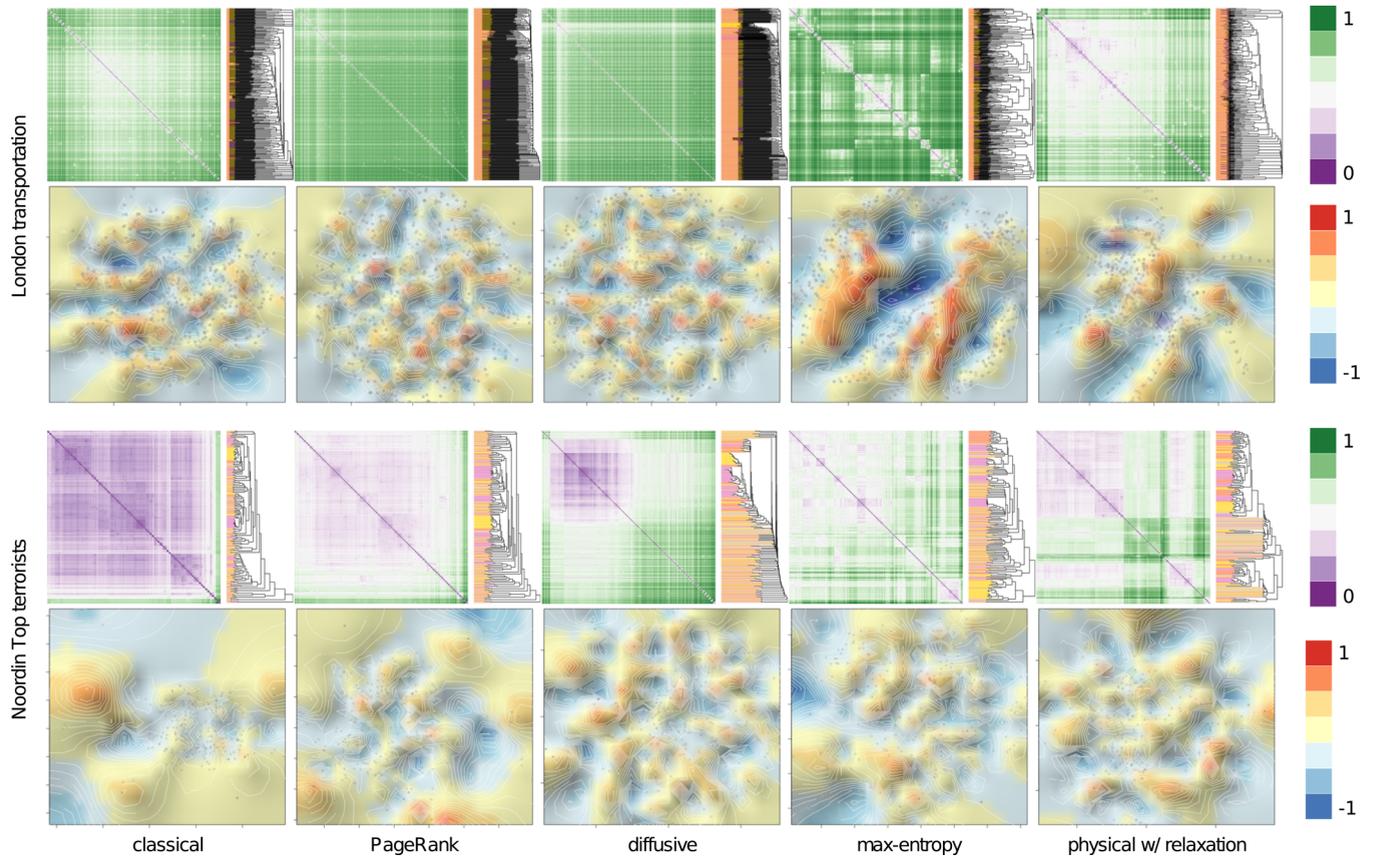
FIG. 8. Average diffusion distance supramatrices $\bar{D}_t$ organized according to their hierarchical clustering as in previous analyses, and projection in $\mathbb{R}^3$ of the corresponding diffusion manifolds, in the case of the two real multiplex networks: the London public transportation network (top panels) and the social relationships of Indonesian terrorists (bottom panels). In both cases, nodes have been embedded in space, more specifically in $[-1, 1]^3$, through multidimensional scaling, a metric-preserving embedding method that depends on two parameters: a distance (or dissimilarity) matrix and an embedding dimension [50] (see Appendix C for details). To facilitate the visualization of the three-dimensional embeddings, we draw the surface which better approximates the cloud of nodes in $\mathbb{R}^3$ and project it on the plane, encoding the third dimension with colors. Nodes are shown as dots on the top of the surfaces.

originally disjoint diffusion manifolds approach each other because of the presence of multilayer information pathways. Our results highlight that also in the regime of partially interconnected multiplex, where not all replicas of a physical node are interacting—which in the real world could mean a failed connection between a bus and a train station—cross-layer pathways form, allowing the interlayer information exchange. Furthermore, as we move toward the fully interconnected multilayer, distances become smaller (as shown by the Frobenius norm), and we highlighted two different behaviors of the manifolds, depending on the meso and macroscale of the single layers: in the Barabasi-Albert model the distinctive mesoscale structure, i.e., the presence of hubs, remains clearly visible, while the community structure in the Girvan-Newman networks is washed out in the diffusion space, above all for small diffusion times.

Finally, we have applied our framework to two empirical multilayer systems, namely, the public transportation of London and the social network of Noordin terrorists. The diffusion geometry corresponding to different random walk dynamics are not necessarily distinct, and we have developed a quantitative method to assess the correlation between the underlying multilayer diffusion manifolds. In the case of the

transportation system, we find that the MERW, which in the synthetic networks was able to separate the layers, highlights two groups of near nodes, that are not captured by other dynamics. This may suggest that (1) the structure of this system, at different scales, has features that are characteristic of different models (e.g., heterogeneity and communities) and (2) different dynamics induce different manifolds and consequently, the analysis of networked systems embedded into space cannot exclude the analysis of the dynamics itself. Conversely, in the case of the social system, we find that the metrics have generally higher correlations, so that their latent diffusion spaces may be likewise similar. The hierarchical structure unveiled by the supradistance matrices seems to suggest a cross-layer core-periphery functional organization.

In this work we extended the mathematical framework for diffusion distances, introduced in [13], to different random walk dynamics and different interconnected layers. We highlighted some interesting results in the sense of functional mesoscale structures. Nonetheless, the thorough analysis of functional multilayer communities is beyond the scope of this work and left for future work. Our work provides a tool for the analysis of multilayer systems from a network geometry

FIG. 9. Estimating the similarity between diffusion manifolds corresponding to different RW dynamics, evaluated on the London multimodal transportation network. We calculate the Pearson's correlation (encoded by size and color) between the entries of pairs of supradistance matrices (Mantel's statistic), which is then tested for significance by permutation (permutation test), with $\alpha = 0.001$. The test is general, because it applies directly on distance matrices, whereas any test performed on the low-dimensional embedding of diffusion manifolds would be less precise because of the information loss during projections.

perspective [10]. Since the latent diffusion geometry is induced by network-driven processes, our framework provides also a complementary view to structural analysis, such as the one provided by hyperbolic network geometry [5,61,62], recently used to analyze multilayer networks [63], and higher-order analysis [64].

FIG. 10. Same as in Fig. 9 for the Noordin Top Terrorists Network.

APPENDIX A: RANDOM WALKS AND MARKOV CHAINS

A Poisson process is a right-continuous process $(X_t)_{t \geqslant 0}$ with values in $\{0, 1, 2, \ldots\}$ and holding times $S_1, S_2, \ldots$ ($S_i = J_i - J_{i-1}$ is the time occurring between the random jump times J_{i-1} and J_i) that are independent exponential random variables of rate $0 < \lambda < \infty$. In a generalized Poisson process (or birth process) the parameter λ is allowed to depend on the current state of the process. Given its *birth rates* $0 \leqslant q_i < \infty$ for $i = 0, 1, 2, \ldots$, $(S_i)_{i \in \mathbb{N}^+}$ are independent exponential random variables with rates q_i. Finally, a continuous-time Markov chain (MC) $(X_t)_{t \geqslant 0}$ on a finite set I with generator Q and initial distribution $\mathbf{p}_0$ can be described in terms of a Poisson process. Each state $i \in I$ of the process is a chamber, and doors close the passage to the other states. From time to time a single door opens (events cannot be simultaneous) allowing the process to change state and the doors open at the jump times of a Poisson process of rate q_{ij} [65]. The *generator* of the MC is indicated by Q, because it is a particular matrix, called Q matrix in [65], satisfying three conditions:

(i) $0 \leqslant -q_{ii} < \infty$

(ii) $q_{ij} \geqslant 0 \quad \forall i \neq j$

(iii) $\sum_j q_{ij} = 0 \quad \forall i$.

To recap, a continuous-time MC can be (equivalently) defined in terms of its jump chain and holding times, or of its transition probabilities given by the solution of the forward equation [65, Theorem 2.8.2].

APPENDIX B: ON THE DIFFUSION DISTANCE(S) OF SINGLE-LAYER NETWORKS

In this Appendix we assume that $G = (V, E)$ is an undirected, weighted, and connected network and we focus on the (single-layer) diffusion distance based on the classical random walk. In this case, we have that the random walk normalized Laplacian $\tilde{\mathbf{L}} = \mathbf{I} - \mathbf{D}^{-1}\mathbf{W} = \mathbf{D}^{-\frac{1}{2}}\mathcal{L}\mathbf{D}^{\frac{1}{2}}$, where $\mathcal{L}^{\mathrm{sym}} = \mathbf{D}^{-\frac{1}{2}}(\mathbf{D} - \mathbf{W})\mathbf{D}^{-\frac{1}{2}}$ is the the symmetric normalized Laplacian matrix [66]. The latter has a spectrum of real, non-negative eigenvalues, and its eigenvectors can be chosen to form an orthogonal matrix $\mathbf{Q}$, so that $\mathcal{L}^{\mathrm{sym}} = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^T$. It follows that $\tilde{\mathbf{L}} = \mathbf{D}^{-\frac{1}{2}}\mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^T\mathbf{D}^{\frac{1}{2}}$ and consequently $e^{-t\tilde{\mathbf{L}}} = \mathbf{D}^{-\frac{1}{2}}\mathbf{Q}e^{-t\boldsymbol{\Lambda}}\mathbf{Q}^T\mathbf{D}^{\frac{1}{2}}$. Let us indicate by $\phi_\ell(i) = Q_j^i$ the ith component of the ℓth eigenvector of $\mathcal{L}^{\mathrm{sym}}$, then

$$(e^{-t\tilde{\mathbf{L}}})_k^i = \sum_{\ell=1}^{N} \frac{\phi^\ell(i)}{\sqrt{s_i}} \phi_\ell(k)\sqrt{s_k}\, e^{-t\lambda_\ell}$$

$$= \sum_{\ell=1}^{N} \psi^\ell(i)\varphi_\ell(k)e^{-t\lambda_\ell}$$

with $\psi^\ell(i) = \frac{\phi^\ell(i)}{\sqrt{s_i}}$ and $\varphi_\ell(k) = \phi_\ell(k)\sqrt{s_k}$ being respectively the ith and kth components of the ℓth right and left eigenvectors of $\tilde{\mathbf{L}}$. We can write the square diffusion distance at a fixed time $t > 0$ between $i, j \in V$ [defined in Eq. (2)] as

$$D_t^2(i, j) = \sum_{k=1}^{N} \left[(e^{-t\tilde{\mathbf{L}}})_k^i - (e^{-t\tilde{\mathbf{L}}})_k^j \right]^2$$

$$= \sum_{k=1}^{N} \left[\sum_{\ell=1}^{N} \frac{\phi^\ell(i)}{\sqrt{s_i}} \phi_\ell(k)\sqrt{s_k}\, e^{-t\lambda_\ell} \right.$$

$$-\sum_{\ell=1}^{N}\frac{\phi^{\ell}(j)}{\sqrt{s_j}}\phi_{\ell}(k)\sqrt{s_k}e^{-t\lambda_{\ell}}\Bigg]^2$$

$$=\sum_{k=1}^{N}s_k\sum_{\ell=1}^{N}e^{-2t\lambda_{\ell}}\left[\frac{\phi^{\ell}(i)}{\sqrt{s_i}}-\frac{\phi^{\ell}(j)}{\sqrt{s_j}}\right]^2. \qquad (B1)$$

Now, exploiting the spectral representation of the exponential Laplacian matrix, we show that for a fixed value of t, D_t is a distance function. Here we assume that $p_t^i(t) = (e^{-t\tilde{\mathbf{L}}})_j^i \neq \pi_j$, that is if and invariant distribution $\boldsymbol{\pi}$ for the Markov chain exists, t is small enough not to be in equilibrium. First, for any pair of nodes i, j, $D_t(i, j) = D_t(j, i)$ and $D_t(i, j) \geqslant 0$. $D_t(i, j) = 0$ if $i = j$. The case $(e^{-t\tilde{\mathbf{L}}})_k^i = (e^{-t\tilde{\mathbf{L}}})_k^j$ for all k is impossible, since it would imply two rows of $e^{-t\tilde{\mathbf{L}}}$ be identical. This, in turn, would imply $\det(e^{-t\tilde{\mathbf{L}}}) = 0$, but $\det(e^{-t\tilde{\mathbf{L}}}) = \det(e^{-t\mathbf{\Lambda}}) = \prod_{\ell}e^{-t\lambda_{\ell}} > 0$. Hence, $D_t(i, j) = 0$ if and only if $i = j$. The triangular inequality follows from the triangular inequality of the norm norm. Finally, D_t is an Euclidean distance [67] and the diffusion distance matrix is an Euclidean distance

matrix. As a matter of fact, D_t is the Euclidean distance in $\mathbb{R}^N$ between the row vectors forming $e^{-t\tilde{\mathbf{L}}}$.

The physical interpretation of the diffusion distance can be found looking at Eq. (2) written as

$$D_t^2(i, j) = \|\mathbf{p}(t|i)\|_2^2 + \|\mathbf{p}(t|j)\|_2^2 - 2\langle\mathbf{p}(t|i), \mathbf{p}(t|j)\rangle$$

$$= \sum_k[p_k(t|i)^2 + p_k(t|j)^2 - 2p_k(t|i)p_k(t|j)].$$

The square terms correspond to the joint probability of two random walkers starting both in the same node to be found at node k at time t (i.e., two independent copies of the same random process), while $p_k(t|i)p_k(t|j)$ corresponds to the probability that they meet in k, starting one in i and the other in j. Since $D_t^2(i, j) \geqslant 0$ we have that $\|\mathbf{p}(t|i)\|_2^2 + \|\mathbf{p}(t|j)\|_2^2 \geqslant 2\langle\mathbf{p}(t|i), \mathbf{p}(t|j)\rangle$, i.e., given a pair of random walkers, the probability that they meet at the same node at time t if they start both at i or both at j is larger than if they start one at i and one at j. If there are many walks connecting i and j then $\sum_k p_k(t|i)p_k(t|j)$ is large and, consequently, $D_t(i, j)$
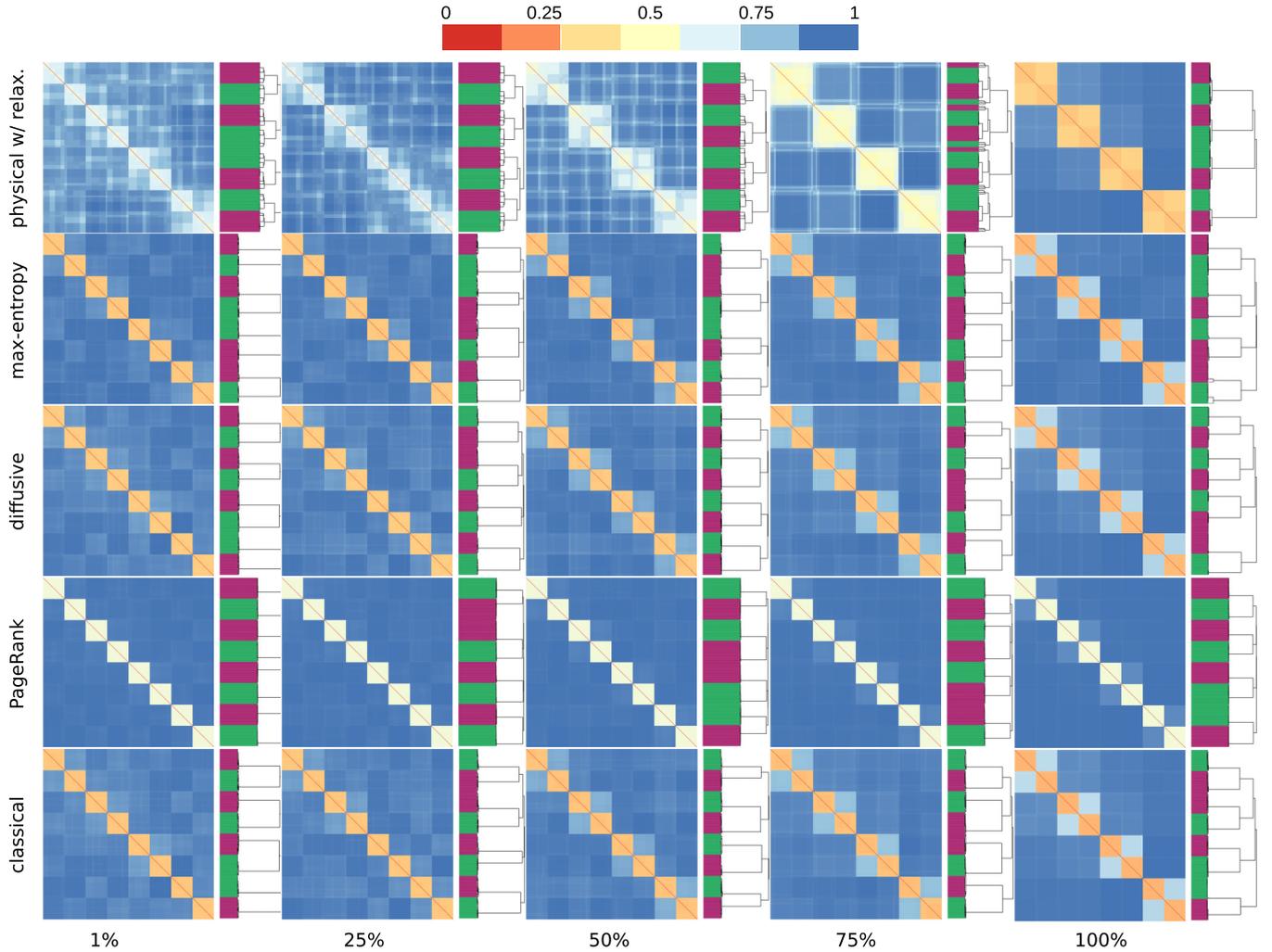

FIG. 11. Average diffusion distance $\bar{D}_t$ for a two-layer multiplex with mesoscale organization in strong communities, partly overlapping across layers. Note that at variance with edge overlapping, here partition overlapping is defined in terms of nodes belonging to the same group without requiring those nodes to be connected by an overlapping edge.

becomes small. we can see that this distance is small if $\langle \mathbf{p}(t|i), \mathbf{p}(t|j) \rangle = \sum_k p_k(t|i) p_k(t|j)$ is large.

Coifman and Lafon, in their paper [33], suggested another interpretation of D_t (based on discrete-time Markov chains) as a distance between bump functions. The authors observed that for a fixed x in a data set X, the transition probability out of x at time t, $p_t(x, \cdot)$, is a bump function centered at x and of width increasing with t. In network terms, $p^i(t)$ is a function defined on V and centered at i; as time increases the random walker explores a larger part of the network, and there is a larger probability to find it far away from the starting node i.

APPENDIX C: MULTIDIMENSIONAL SCALING

Multidimensional scaling (MDS) is a collection of methods aiming to find structure in high-dimensional data, given a $n \times n$ dissimilarity matrix Δ between the observations and the dimension p of the Euclidean space in which data are embedded [50]. The final goal is to find a space configuration of the data points in $\mathbb{R}^p \ni \mathbf{x}_1, \ldots, \mathbf{x}_n$, such that $\|\mathbf{x}_i - \mathbf{x}_j\| \approx \delta_{ij}$, where δ_{ij} are the original distances and $\| \cdot \|$ is the Euclidean distance in $\mathbb{R}^p$.

The MDS problem can be solved analytically or using iterative procedures, like the majorization algorithm. The first case, involves a double-centering step of the dissimilarity matrix and then its eigendecomposition. The second, involves the minimization of a loss function, in particular, Kruskal's *stress-1*. Given a configuration X of N points in $\mathbb{R}^p$ with $1 \leqslant p \leqslant N - 1$:

$$\sigma(X) = \sum_i \sum_j [\delta_{ij} - d_{ij}(X)]^2. \qquad \text{(C1)}$$

If Δ is not a proper distance matrix, δ_{ij} are called dissimilarities, and they may need to be transformed into $\hat{d}_{ij} = f(\delta_{ij})$ before the scaling, for instance, through a strictly increasing monotonic function f such that order is preserved $\delta_{ij} < \delta_{kl} \Rightarrow f(\delta_{ij}) < f(\delta_{kl})$.

In this work we use the second approach (MDS solved through majorization) applied to distance or supradistance matrices and with embedding dimension $p = 3$. The R package and function used is SMACOF::MDS [68].

APPENDIX D: DIFFUSION DISTANCE AND COMMUNITY OVERLAP

Figure 11 shows how the diffusion geometry of networks with strong mesoscale organization is influenced by layer-layer correlations, in particular community overlap. As the overlap between the communities in the two layers grows, macrocommunities across layers emerge.

[1] C. Song, S. Havlin, and H. A. Makse, Nature (London) **433**, 392 (2005).

[2] C. Song, S. Havlin, and H. A. Makse, Nat. Phys. **2**, 275 (2006).

[3] M. A. Serrano, D. Krioukov, and M. Boguñá, Phys. Rev. Lett. **100**, 078701 (2008).

[4] D. Taylor, F. Klimm, H. A. Harrington, M. Kramár, K. Mischaikow, M. A. Porter, and P. J. Mucha, Nat. Commun. **6**, 7723 (2015).

[5] F. Papadopoulos, M. Kitsak, M. Á. Serrano, M. Boguñá, and D. Krioukov, Nature (London) **489**, 537 (2012).

[6] A. Allard and M. Á. Serrano, PLoS Comput. Biol. **16**, e1007584 (2020).

[7] M. Boguñá, D. Krioukov, and K. C. Claffy, Nat. Phys. **5**, 74 (2009).

[8] M. Xu, Q. Pan, A. Muscoloni, H. Xia, and C. V. Cannistraci, Nat. Commun. **11**, 2849 (2020).

[9] M. Boguñá, F. Papadopoulos, and D. Krioukov, Nat. Commun. **1**, 62 (2010).

[10] M. Boguñá, I. Bonamassa, M. De Domenico, S. Havlin, D. Krioukov, and M. Á. Serrano, Nat. Rev. Phys. **3**, 114 (2021).

[11] E. Estrada, Linear Algebra Appl. **436**, 4317 (2012).

[12] D. Brockmann and D. Helbing, Science **342**, 1337 (2013).

[13] M. De Domenico, Phys. Rev. Lett. **118**, 168301 (2017).

[14] C. Hens, U. Harush, S. Haber, R. Cohen, and B. Barzel, Nat. Phys. **15**, 403 (2019).

[15] M. Kac, Am. Math. Mon. **54**, 369 (1947).

[16] J.-P. Bouchaud and A. Georges, Phys. Rep. **195**, 127 (1990).

[17] S. Havlin and D. Ben-Avraham, Adv. Phys. **51**, 187 (2002).

[18] J. D. Whitfield, C. A. Rodríguez-Rosario, and A. Aspuru-Guzik, Phys. Rev. A **81**, 022323 (2010).

[19] S. E. Venegas-Andraca, Quant. Info. Proc. **11**, 1015 (2012).

[20] V. Zaburdaev, S. Denisov, and J. Klafter, Rev. Mod. Phys. **87**, 483 (2015).

[21] L. Giuggioli, Phys. Rev. X **10**, 021045 (2020).

[22] N. Masuda, M. A. Porter, and R. Lambiotte, Phys. Rep. **716–717**, 1 (2017).

[23] M. Rosvall and C. T. Bergstrom, Proc. Natl. Acad. Sci. USA **105**, 1118 (2008).

[24] J.-C. Delvenne, S. N. Yaliraki, and M. Barahona, Proc. Natl. Acad. Sci. USA **107**, 12755 (2010).

[25] M. Faccin, P. Migdał, T. H. Johnson, V. Bergholm, and J. D. Biamonte, Phys. Rev. X **4**, 041012 (2014).

[26] R. Lambiotte, J.-C. Delvenne, and M. Barahona, IEEE Trans. Netw. Sci. Eng. **1**, 76 (2014).

[27] M. De Domenico, A. Lancichinetti, A. Arenas, and M. Rosvall, Phys. Rev. X **5**, 011027 (2015).

[28] Z. Burda, J. Duda, J.-M. Luck, and B. Waclaw, Phys. Rev. Lett. **102**, 160602 (2009).

[29] M. De Domenico, A. Solé-Ribalta, S. Gómez, and A. Arenas, Proc. Natl. Acad. Sci. USA **111**, 8351 (2014).

[30] S. Brin and L. Page, Comput. Netw. ISDN Syst. **30**, 107 (1998).

[31] M. De Domenico and J. Biamonte, Phys. Rev. X **6**, 041062 (2016).

[32] A. Ghavasieh and M. De Domenico, Phys. Rev. Res. **2**, 013155 (2020).

[33] R. R. Coifman and S. Lafon, Appl. Comput. Harmonic Anal. **21**, 5 (2006).

[34] M. Cao, H. Zhang, J. Park, N. M. Daniels, M. E. Crovella, L. J. Cowen, and B. Hescott, PLoS ONE **8**, e76339 (2013).

[35] D. K. Hammond, Y. Gur, and C. R. Johnson, in *2013 IEEE Global Conference on Signal and Information Processing* (IEEE, New York, 2013), pp. 419–422.

[36] G. Bertagnolli, C. Agostinelli, and M. De Domenico, J. Complex Netw. **8**, cnz041 (2019).

[37] M. De Domenico, A. Solé-Ribalta, E. Cozzo, M. Kivelä, Y. Moreno, M. A. Porter, S. Gómez, and A. Arenas, Phys. Rev. X **3**, 041022 (2013).

[38] M. Kivelä, A. Arenas, M. Barthélemy, J. P. Gleeson, Y. Moreno, M. A. Porter, and E. Estrada, J. Complex Netw. **2**, 203 (2014).

[39] M. De Domenico, C. Granell, M. A. Porter, and A. Arenas, Nat. Phys. **12**, 901 (2016).

[40] S. Gomez, A. Diaz-Guilera, J. Gomez-Gardenes, C. J. Perez-Vicente, Y. Moreno, and A. Arenas, Phys. Rev. Lett. **110**, 028701 (2013).

[41] S. Havlin, R. Blumberg Selinger, M. Schwartz, H. E. Stanley, and A. Bunde, Phys. Rev. Lett. **61**, 1438 (1988).

[42] J. Gillis, in *Mathematical Proceedings of the Cambridge Philosophical Society*, Vol. 51 (Cambridge University Press, Cambridge, 1955), pp. 639–651.

[43] F. Battiston, V. Nicosia, and V. Latora, New J. Phys. **18**, 043035 (2016).

[44] M. De Domenico, A. Solé-Ribalta, E. Omodei, S. Gómez, and A. Arenas, Nat. Commun. **6**, 6868 (2015).

[45] T. G. Kolda and B. W. Bader, SIAM Rev. **51**, 455 (2009).

[46] R Core Team, *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, 2019).

[47] D. Edler, L. Bohlin, and M. Rosvall, Algorithms **10**, 112 (2017).

[48] J. I. Perotti, N. Almeira, and F. Saracco, Phys. Rev. E **101**, 062148 (2020).

[49] D. J. Klein and M. Randić, J. Math. Chem. **12**, 81 (1993).

[50] T. F. Cox and M. A. Cox, *Multidimensional Scaling* (Chapman and Hall/CRC, Boca Raton, FL, 2000).

[51] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).

[52] M. De Domenico, V. Nicosia, A. Arenas, and V. Latora, Nat. Commun. **6**, 6864 (2015).

[53] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).

[54] M. Girvan and M. E. J. Newman, Proc. Natl. Acad. Sci. USA **99**, 7821 (2002).

[55] N. Roberts and S. F. Everton, Terrorist data: Noordin top terrorist network (subset) [machine-readable data file] (2011), https://sites.google.com/site/sfeverton18/research/appendix-1.

[56] F. Battiston, V. Nicosia, and V. Latora, Phys. Rev. E **89**, 032804 (2014).

[57] N. Mantel, Cancer Res. **27**, 209 (1967).

[58] P. Legendre and L. F. J. Legendre (ed.), in *Numerical Ecology*, Developments in Environmental Modelling, Vol. 24 (Elsevier, Oxford, UK, 2012), pp. 521–624.

[59] X. Zhang, S. Boccaletti, S. Guan, and Z. Liu, Phys. Rev. Lett. **114**, 038701 (2015).

[60] G. F. de Arruda, G. Petri, and Y. Moreno, Phys. Rev. Res. **2**, 023032 (2020).

[61] D. Krioukov, F. Papadopoulos, M. Kitsak, A. Vahdat, and M. Boguñá, Phys. Rev. E **82**, 036106 (2010).

[62] G. Bianconi and C. Rahmede, Sci. Rep. **7**, 41974 (2017).

[63] K.-K. Kleineberg, M. Boguñá, M. Á. Serrano, and F. Papadopoulos, Nat. Phys. **12**, 1076 (2016).

[64] F. Battiston, G. Cencetti, I. Iacopini, V. Latora, M. Lucas, A. Patania, J.-G. Young, and G. Petri, Phys. Rep. **874**, 1 (2020).

[65] J. R. Norris, *Markov Chains*, Cambridge Series in Statistical and Probabilistic Mathematics (Cambridge University Press, Cambridge, 1998).

[66] F. R. Chung and F. C. Graham, *Spectral Graph Theory*, CMBS Regional Conference Series in Mathematics, 92 (American Mathematical Society, Providence, RI, 1997).

[67] L. Liberti, C. Lavor, N. Maculan, and A. Mucherino, SIAM Rev. **56**, 3 (2014).

[68] J. de Leeuw and P. Mair, J. Stat. Softw. **31**, 1 (2009).