Enhancing transport properties in interconnected systems without altering their structure

Arsham Ghavasieh[1,2] and Manlio De Domenico[2,*]

[1]*Universitá degli Studi di Trento, Via Sommarive 5, 38123 Povo (TN), Italy*
[2]*Fondazione Bruno Kessler, Via Sommarive 18, 38123 Povo (TN), Italy*

Units of complex systems—such as neurons in the brain or individuals in societies—must communicate efficiently to function properly: e.g., allowing electrochemical signals to travel quickly among functionally connected neuronal areas in the human brain, or allowing for fast navigation of humans and goods in complex transportation landscapes. The coexistence of different types of relationships among the units, entailing a multilayer representation in which types are considered as networks encoded by *layer*s, plays an important role in the quality of information exchange among them. While altering the structure of such systems—e.g., by physically adding (or removing) units, connections, or layers—might be costly, coupling the dynamics of subset(s) of layers in a way that reduces the number of redundant diffusion pathways across the multilayer system, can potentially accelerate the overall information flow. To this aim, we introduce a framework for *functional reducibility* which allow us to enhance transport phenomena in multilayer systems by coupling layers together with respect to dynamics rather than structure. Mathematically, the optimal configuration is obtained by maximizing the deviation of system's entropy from the limit of free and noninteracting layers. Our results provide a transparent procedure to reduce diffusion time and optimize noncompact search processes in empirical multilayer systems, without the cost of altering the underlying structure.

I. INTRODUCTION

A wide variety of social, natural and artificial systems are inherently complex [1]. For instance, societies exhibit rich microscopic dynamics, at the level of single individuals, that might lead to emergent collective phenomena at larger scales [2]—e.g., financial collapses or revolutions—which are usually difficult to predict [3]. Natural systems, like ecological ones, are characterized by complex topologies that, in presence of interdependencies with other networks, exhibit a richer response to perturbations with respect to the case where they are considered in isolation [4]. Similar phenomena have been observed also for engineering systems, from transportation to communication networks, where their resilience to targeted attacks or random failures of their units is important for applications [5–7], such as robustness enhancement or recovery strategies [8,9].

In fact, complex systems are characterized by a wide variety of physical attributes and dynamics, making difficult to analyze them within a unified framework. What do electrochemical signals exchanged among neurons in the human brain, mobility of goods/people/vehicles between different areas of a urban ecosystem, spreading of an infectious pathogen through social contact patterns, and financial transactions

through the backbone of a stock market have in common? From a very general perspective, the corresponding systems consist of interconnected units exchanging *information* similarly to how computers and servers exchange packets of bits through the Internet. In fact, it has been originally conjectured by Murray Gell-Mann that despite their differences, complex systems resemble one another in the way they handle information and, consequently, studying how information is exchanged and processed provides a promising starting point for exploring transport phenomena in order to understand how such systems operate.

A general way to model the propagation of information through complex structures is by means of diffusive processes, such as random walks [10], since they are versatile models that allows one to cope with the uncertainty in the structure—e.g., temporary link failures or damages—of complex systems, they are based on local knowledge of structure [11]—which is usually less expensive than models relying only on the global knowledge of the topology, like shortest-path-based ones. Nevertheless, there is increasing evidence that units of real complex systems do not preferentially exchange information along shortest paths, too expensive in terms of routing. In fact, the human brain operates in an intermediate communication regime, neither based on shortest-path routing nor on a diffusion process [12], while the Internet relies on routing tables—more specifically, forwarding tables chosen by the routing algorithm—on each machine to find suboptimal communication pathways through the routing hierarchy [13,14]. Even individual human mobility, usually assumed to be far from random [15] is affected by arbitrariness of individuals' actions responsible for a stochastic component [16] in trip patterns characterizing human flows [17], which

---

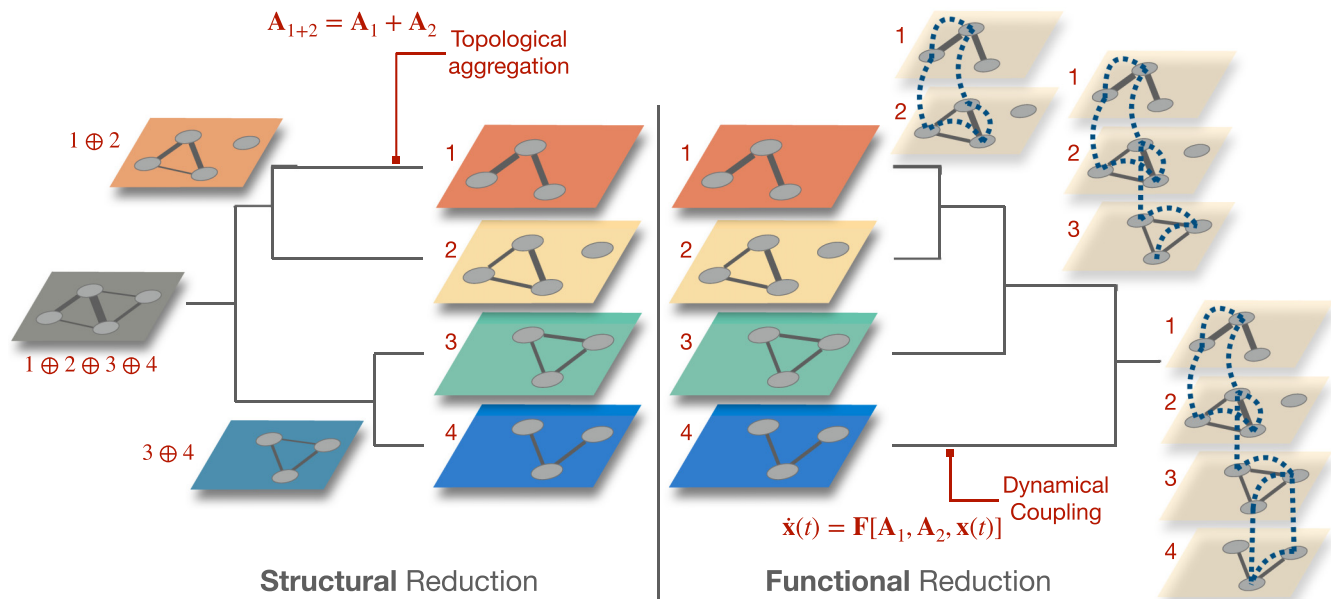*Corresponding author: mdedomenico@fbk.eu

FIG. 1. Structural vs functional reducibility of multilayer systems. In structural reduction (left), one alters the structure by iteratively aggregating layers, e.g., by summing the corresponding adjacency matrices. However, this type of intervention usually comes with some costs (e.g., temporal, infrastructural, economic, etc.). Here, we propose functional reduction (right), an alternative approach where the structure of the system is not altered while coupling the dynamics on the top of layers. While any dynamics is allowed in principle, diffusive processes are particularly suitable ones because they allow for theoretical and computational treatment. In this work, random walk dynamics is considered and layers are functionally coupled if their network states are not distinguishable from the point of view of a walker (i.e., the same color is assigned to functionally coupled layers). From an application perspective, this approach corresponds to shared tickets for public multimodal transport infrastructure or shared promotions for users who are part of multiple social networks at the same time.

is usually modeled by means of stochastic processes ranging from random walks to variants of brownian motion and Levy flights [18]. In this work, we present a way to alter the dynamics on top of multilayer networks, while keeping their structure (see Fig. 1). Without changing the underlying structure, shortest paths across the whole multiplex system do not change accordingly. This unfortunate feature of shortest paths makes them useless to enhance transport properties under the hard constraint of keeping the structure unaltered.

Therefore constituents of complex systems tend to or are designed for exchanging information in a very efficient way to function properly. An emblematic example is given by the human brain, where the collective dynamics of billions of neurons is responsible for coordinating the human body and cognitive activities, while keeping energy cost as low as possible [19]. However, a deeper understanding of how to act on the system to enhance or hinder transport phenomena—such as navigability [20]—continues to elude us, because they depend on the interplay between structure and dynamics of a complex system, which is usually difficult to model and quantify [21]. Nevertheless, understanding such an interplay represents one of the most important challenges in complexity science, with promising advances. In fact, it has been recently shown that the relation between structure and dynamics of information flow can be unraveled by analyzing how perturbations propagate through the system [22]. The complex interplay between structure and dynamics of information propagation has been also mapped to its latent geometry to better

understand network-driven contagion phenomena [23] and, more recently, it has been shown how signal propagation is able to capture the role of network connectivity in propagating local information, thus linking the topology to the observed spatiotemporal spread of perturbative signals across it [24].

However, the same task is even more challenging when the system of interest can be described in terms of a multilayer network [25–31]. For instance, multilayer models for urban and regional transportation systems—accounting for different means, from rails, to ships and flights, serving the same geographical areas simultaneously—have highlighted that an efficient flow might be hindered by the lack of synchronization between different layers [32].

Up to date, it is still unknown how to enhance flow distribution in multilayer systems, especially under constraints. For instance, one might want to add new fast connections (e.g., highways, tube, flights, etc.) between distant geographic areas to speed up the traffic flow. However, since each connection comes with an economic cost, it is very unlikely that most of them can be realized in practice. Similarly, one might avoid to cut existing connections, even if desirable in some cases, because they might have a high societal cost in terms of impact on the population. It has been recently shown that for multilayer networks with interlinks, changing the weights of interlinks can enhance the diffusion on top of the networks [27] and lead to the phenomenon of "super diffusion." However, in many complex systems, as the ones which can be modeled by coupled networks, interlinks either do not exist or it is not trivial to assign weights to them. Therefore the challenge is

to enhance transport properties without altering the existing structure: this can be achieved, in principle, by functionally coupling together layers with similar flow patterns. When each layer of the multilayer system is represented by a specific color, an emblematic way to understand functional coupling of two layers is to assign the same color to both of them and make them indistinguishable, as illustrated in Fig. 1.

It is worth remarking that this approach leaves the structures of layers intact, while only altering the dynamics on top of them, in a way that the dynamical process can not recognize functionally coupled layers as distinct ones. For instance, this type of solutions is adopted, to some extent, by airlines proposing shared flights to their customers.

Intuitively, one might think that reducing structural redundancy—defined in terms of replicated connections across layers—can provide a solution. Remarkably, the recent analysis of common estimators of structural redundancy, such as edge overlapping, revealed that the presence of redundant connectivity might boost the robustness of multiplex networks [33]. However, when more complex indicators of redundancy are needed, our understanding is dramatically more limited. On the one hand, recent studies focused on quantifying the distance between two complex networks in terms of their structural dissimilarities [34]—as a distance between probability distributions extracted from the networks—or their spectral information content [35] have been successfully used to compare classical, single-layer, networks. On the other hand, structural dissimilarities [36] and spectral entropy divergence [37] between layers have been recently proposed as effective approaches for reducing the structure of such systems, based on different ordering strategies. However, methods relying only on structural features are not suitable for our purposes—as they are costly in practice—and they perform poorly, as we show later. The reason is that they are based on heuristics, lacking a fundamental understanding of the underlying physics.

We show that a fundamentally different perspective must be considered instead of structural redundancy: by using random walk dynamics as a proxy for information flow, transport phenomena are enhanced when subsets of layers are functionally grouped together in such a way that the corresponding diffusion pathways are maximally different. Therefore one must analyze the functional redundancy of a multiplex system, defined by abundant and redundant diffusion pathways that distribute the flow between different system's units.

In the following, for sake of clarity, to avoid any confusion with structural redundancy, we will refer to functional reduction and functionally reduced networks to indicate a loss of redundancy in diffusion pathways across layers. At variance with structural reducibility, functionally reducible multiplex systems consists of layers that can be functionally grouped into subsets to enhance transport, whereas irreducible systems do not allow for any functional aggregation. The difference between structural reducibility that alters the structure of multilayer networks, and the proposed functional reducibility that only alters the dynamics on top of multilayer networks, is illustrated in Fig. 1.

In this work, we develop tools familiar to physicists—such as partition function and mean-field calculations—to investigate the interplay between structure and dynamics of

multilayer systems from a spectral perspective—an approach revealing successful for a similar purpose in the case of classical networks [38–41]. By exploiting the relation between the flow distribution in a multiplex network and the spectral diversity of its layers, our framework can provide a broad spectrum of applications—from more optimal supply strategy that combine transportation of goods to more efficient navigability of urban areas by allowing for multimodal trips with the same ticket—that can be achieved by devising tailored policies targeting the dynamics of systems, without the cost of changing their structure.

II. STATISTICAL PHYSICS OF RANDOM WALKS IN COMPLEX NETWORKS

In this section, we introduce the state of multiplex networks, discuss the role of their structure in hindering the information flow, find an inequality relating partition functions of single layers to the partition function of the whole system and develop a mean-field approach to allow for analytical derivations of the next sections. Note that most mathematical details are provided in Appendices, while we provide here the theoretical grounds required for defining and understanding the functional reducibility of multilayer systems.

A. Information flow in multiplex networks

Multiplex networks are a special class of multilayer systems providing a successful and widely used model for a broad variety of empirical systems [29,30]. This class of networks is characterized by nodes with multiple types of relationships or interactions, simultaneously, which are encoded by layers. Here, physical nodes correspond to the physical units of the system, whereas state nodes correspond to the replicas of a physical node across layers.

Information flow in complex networks has been successfully modeled by diffusive processes such as random walks [10]. However, while random walk dynamics has been previously introduced for interconnected multilayer systems [6,26], the case of multiplex systems has been recently studied only by using the full aggregation of the system into a single-layer network [42]. Clearly, the dynamics of a random walker on the aggregated representation of a multiplex system can not coincide with the multiplex dynamics.

To describe the random walk dynamics on top of multiplex networks, we first find the corresponding transition matrix, which is given by $\mathbf{T} = \langle \mathbf{T}^{(\ell)} \rangle$ (see Appendix B), and encodes the probability of jumps from each node to any other. Similarly, the normalized Laplacian matrix corresponding to the multiplex random walk dynamics can be obtained as $\mathcal{L} = \mathbf{I} - \mathbf{T} = \langle \mathcal{L}^{(\ell)} \rangle$, being $\mathbf{I}$ the identity matrix. This means that the normalized Laplacian matrix encoding random walks on a multiplex network is the weighted average of Laplacian matrices of layers. Importantly, each layer's weight corresponds to its contribution to the overall flow of information. As these weights are often hard to assess or not known at all, when this is the case we assume that there is no preference on layers, and from now on we consider uniform distribution of weights, $1/L$, for a multiplex network of L layers. Nevertheless, it is worth remarking that the proposed method and the

theoretical framework are valid for any general distributions of weights.

The propagator (see Appendix B) governs the temporal evolution of the random walk and, for this reason, it encodes information about the interplay between the structure of the system and the random search dynamics on the top of it. In fact, it has been recently used to define the state of the system [35], similar in spirit to the approach widely used in quantum statistical mechanics to define the mixed state of entangled quantum units. In this framework, the network state is defined by $\boldsymbol{\rho}(\tau) = \frac{e^{-\tau\mathcal{L}}}{Z}$, where $Z(\tau) = \text{Tr}(e^{-\tau\mathcal{L}})$ is a normalization factor, playing the same role of the partition function of the system. This state, which is formally identical to a density matrix [43], is then used to calculate the Von Neumann entropy of a complex network as $S(\tau) = -\text{Tr}(\boldsymbol{\rho}(\tau)\log_2\boldsymbol{\rho}(\tau))$. See Appendix A for further details.

B. Physical meaning of the partition function

Here, we provide first a deeper understanding of the meaning of the partition function. In fact, $Z(\tau) = N\mathcal{R}(\tau)$, being $\mathcal{R}(\tau) = N^{-1}\sum_{i=1}^{N}e^{-\tau\lambda_i}$ the average return probability of random walker and λ_i is the ith eigenvalue of $\mathcal{L}$ (see Appendix C). As mentioned before, random walk can be used as a proxy of information transport in complex networks. Intuitively, high average return probability is associated with the tendency of structure to trap the information flow in its initial place. So, this correspondence between partition function of the system and average return probability, unravels how partition function relates to transport phenomena in complex networks. Furthermore, one can define density matrices in terms of any valid Laplacian encoding other types of dynamics (e.g., continuous diffusion). In this case, while the average return probability can not even be defined, the partition function still exists and corresponds to the amount of information that is trapped in its initial place and flow more difficultly towards peripheral parts of the network.

As mentioned above, high average return probability indicates that the random walker can not explore efficiently the rest of the network. Consequently, the presence of structural symmetries and abundance of redundant diffusion pathways for information to propagate between system's units are expected. For instance, in an ordered and symmetric structure like a regular lattice, where each node interacts only with its first neighbors and no long-range interactions are possible, the information exchange between distant nodes is slow and inefficient. Conversely, low average return probability indicates that the random walker can navigate the network faster. For instance, breaking down structural order by adding long range interactions between distant nodes, generating topological shortcuts and other defects like in small-world [44] and scale-free [45] networks, information flows faster and transport properties are enhanced, accordingly. $Z(\tau)$, as a measure of transport that we name average dynamical trapping in the following, is able to encode these phenomena. The same arguments apply to multiplex networks, with the difference that in this case we are wondering if the redundancy of diffusion pathways across layers can be reduced by superimposing subsets of layers, in order to reduce the average dynamical trapping.

C. Fundamental inequality for multiplex systems

Multiplexity of interactions among units generate nontrivial dynamical correlations between layers that have no counterpart when layers are considered in isolation. This important difference can be characterized in terms of differences in information content, quantified by average entropy distance of the multiplex and its layers (see Appendix D), that we call *intertwining*. In the next sections, we discuss the central role of intertwining in functional reducibility.

Directly from intertwining, a fundamental inequality between the partition function of a multiplex system as whole and the partition functions of its layers can be derived. This inequality is important, as it relates the transport phenomena of multiplex system and layers, through average dynamical trapping (see Appendix D for details):

$$Z(\tau) \leqslant \prod_{\ell=1}^{L} Z^{(\ell)}(\tau)^{1/L}, \tag{1}$$

where equality holds if and only if all the layers are the same, i.e., the random exploration of the network does not depend on any specific layer of the system. This equality defines the noninteracting scenario where layer-layer correlations do not alter the underlying dynamics: the entropy $S^{(\ell)}(\tau)$ of each layer is calculated separately and the overall entropy is given by their average $S_{\text{nint}}(\tau) = \langle S^{(\ell)}(\tau)\rangle$. Conversely, any topological alteration of the noninteracting scenario introduces a dynamical correlation between layers, requiring the exploration of layers to gather more information about the system: in this case the network consists of interacting layers, where the diffusion dynamics on the whole multiplex network is considered to measure the entropy $S_{\text{int}}(\tau)$.

D. Mean-field entropy

An analytical treatment of Von Neumann entropy is, in general, difficult and we have developed a spectral mean-field theory to cope with this complexity (Appendices E and F):

$$S_{\text{MF}}(\tau) = \frac{1}{\log 2}\left(\tau\frac{Z(\tau)-1}{Z(\tau)} + \log Z(\tau)\right). \tag{2}$$

When the base of logarithm is not specified, we refer to the natural logarithm. In the next section, we use mean-field entropy to find a simple representation of intertwining, in terms of S_{int} and S_{nint}.

III. QUANTIFYING LAYER-LAYER INTERACTIONS: THE INTERTWINING

Using fundamental inequality (1) and mean-field entropy (2), it can be shown that $S_{\text{int}}(\tau) \leqslant S_{\text{nint}}(\tau)$ (Appendix G): i.e., layer-layer interactions lower system's entropy and, consequently, its partition function $Z(\tau)$. We exploit the average information divergence of layers from the whole [see Eq. (D2)] rescaled by its upper bound, to provide a normalized measure that quantifies the importance of describing the system in terms of coupled networks rather than networks in isolation. For values of τ sufficiently large and in absence of isolated state nodes, it can be shown that the relative intertwining reduces to the deviation between interacting and noninteracting

entropies (see Appendix I for details):

$$\mathcal{I}^*(\tau) = 1 - \frac{S_{\text{int}}(\tau)}{S_{\text{nint}}(\tau)}, \tag{3}$$

which is bounded between 0 (i.e., the system is reducible to a single network) and 1 (i.e., the system is irreducible).

Before presenting how to use the relative intertwining for the analysis of multiplex networks, we derive analytically its direct dependence on the partition function and the eigenvalues of the normalized Laplacian matrix, especially on λ_2—the second smallest one—which governs many transport phenomena, as shown in the following. From the master equation, the Laplacian matrix of the multiplex network is given by $\mathcal{L} = \langle \mathcal{L}^{(\ell)} \rangle$: here, we write the Laplacian matrix $\mathcal{L}^{(\ell)}$ of each layer as the perturbation $\mathcal{L}^{(\ell)} = \mathcal{L} + \Delta \mathcal{L}^{(\ell)}$, reflecting in the corresponding eigenvalues as $\lambda_i^{(\ell)} = \lambda_i + \Delta \lambda_i^{(\ell)}$ ($i = 0, 1, \dots, N$).

It can be shown that $\langle \Delta \lambda^{(\ell)} \rangle = 0$ (see Appendix I) and that $\langle (\Delta \lambda^{(\ell)})^2 \rangle \geqslant 0$, the latter quantifying the influence of the perturbation to each layer. The variance is zero only in the case of identical layers.

A quantity of interest is the variance averaged across all layers, i.e., $\overline{\langle (\Delta \lambda^{(\ell)})^2 \rangle}$, which measures the overall influence of the perturbation: it is expected to increase for increasing deviation of layers from the average, i.e., for increasing diversity of the layers. In fact, we show that the relative intertwining is directly proportional to this variance (Appendix I):

$$\mathcal{I}^\star(\tau) \approx \frac{\tau^2}{2} \overline{\langle (\Delta \lambda^{(\ell)})^2 \rangle}, \tag{4}$$

which increases as the diversity of diffusion pathways of layers increases.

Similarly, $Z^{(\ell)}(\tau) = Z_{\text{int}}(\tau) + \Delta Z^{(\ell)}(\tau)$ holds for partition functions and we show that (see Appendix I for details)

$$\mathcal{I}^\star(\tau) \approx \frac{\overline{\Delta Z^{(\ell)}(\tau)}}{Z_{\text{int}}(\tau) - 1}; \; \overline{\Delta Z^{(\ell)}(\tau)} = \frac{1}{L} \sum_{\ell=1}^{L} \Delta Z^{(\ell)}(\tau). \tag{5}$$

Equations (4) and (5) provide the first fundamental result of this work: they show that by minimizing the partition function of the system one maximizes the relative intertwining while favoring the maximum functional diversity of layers. The second fundamental result of this work is the direct relationship between the relative intertwining and the second smallest eigenvalue λ_2 of the normalized Laplacian matrix (see Appendix M for details):

$$\mathcal{I}^\star(\tau) \approx \tau \left(\lambda_2 - \overline{\lambda_2^{(\ell)}} \right). \tag{6}$$

This relation is important because it highlights how maximizing intertwining corresponds to maximize λ_2, which in turn is equal to the inverse of diffusion time and it plays a crucial role for navigability as two important transport properties of complex networks (see Appendix L).

Thus, by maximizing the relative intertwining, one actually enhances the most important transport phenomena on multiplex systems and, as a byproduct, reduces their dimensionality. To use our framework for practical applications, we need a strategy for identifying similar layers in a multiplex network, aggregating them accordingly, and a stopping cri-

terion. We use a novel approach, namely collective cosine distance, which is computationally faster and more reliable than existing methods when dealing with isolated nodes and multiple connected components (see Appendix J) to cluster layers. Note that we perform an exhaustive pairwise search among possible functional groups, *de facto* getting rid of heuristics such as hierarchical clustering [35,37].

IV. FUNCTIONAL REDUCIBILITY OF MULTIPLEX SYSTEMS

For each value of the time τ, we calculate all pairwise distances between layers to to find the most similar ones: at each step m of this sequence, the corresponding value $\mathcal{I}^\star(\tau; m)$ of the relative intertwining is calculated. Finally, the values of relative intertwining are averaged over τ as $\mathcal{I}^\star(m) = \langle \mathcal{I}^\star(\tau; m) \rangle$ and used for comparison against navigability, diffusion time at each aggregation step. We repeat this procedure until the original system is aggregated into a single network, representing the superposition of the whole system from the dynamical perspective. The average relative intertwining is expected to reach a maximum value if any desirable superposition of subsets of layers exists (see Appendix I analytical details): in this case, the resulting system is characterized by enhanced transport properties. At this optimal grouping, in the sense that it is the most desirable configuration among the possible ones, the remaining number of layers can be indicated by L_{opt} and one can define the *reducibility* of a multilayer network as $\chi = (L - L_{\text{opt}})/(L - 1)$ to characterize this system's feature with one scalar. Note that χ quantifies the fraction of layers grouped together: one can use either structural reducibility [37]—maximizing the topological distinguishability from the aggregate—or functional reducibility—maximizing differences in diffusion pathways across layers—defined in this work. However, when transport phenomena are a concern, functional reducibility must be preferred.

A. Synthetic systems

We have performed several numerical experiments to validate our theory on toy models with functionally reducible or irreducible systems. We show one representative case for each case in Fig. 2. For the reducible case, see Figs. 2(a)–2(c), our framework is able to detect functional redundancy and reduce it as expected, while lowering diffusion time, average return probability and dynamical trapping, and increasing navigability. For the irreducible case, see Figs. 2(d)–2(f), a multiplex system consisting of noisy and independent layers is not functionally reducible, as expected. For further synthetic systems, see Appendix N.

Remarkably, all transport properties considered here are enhanced by functionally coupling layers according to our theoretical framework.

In the following, we apply our framework to analyze the functional reducibility of empirical multilayer systems, including biological, social, and transportation ones.

B. Biological system

As a representative biological system, we consider the largest connected component of molecular interactions of
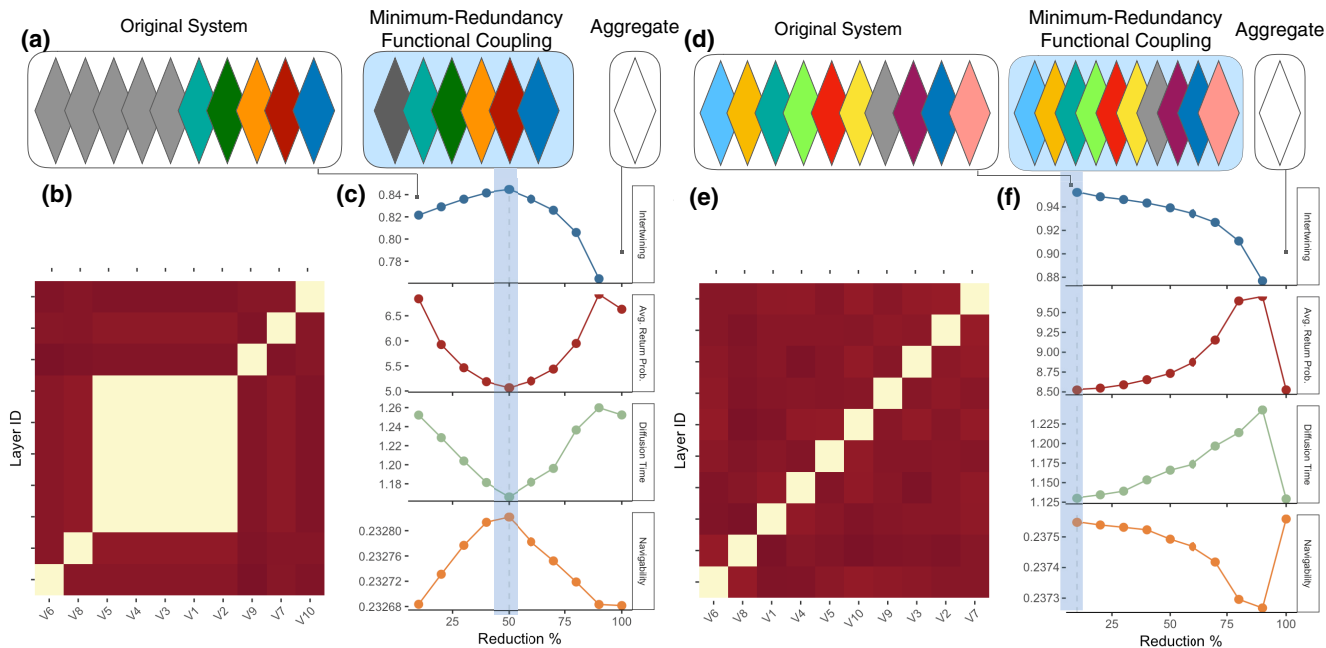
FIG. 2. Functional reducibility in synthetic systems. (a) A multiplex network with 100 nodes, consisting of five identical layers and five distinct random networks with wiring probability 0.2 is illustrated in its simulated form (original system), its form after functionally coupling its layers according to the framework presented in this study (minimum-redundancy functional coupling) and the functional form corresponding to the coupling of all layers simultaneously (Aggregate). (b) Heat-map encoding pairwise dissimilarity (the darker the color, the larger the distance) between layers shown in (a), according to collective cosine distance (see Appendix J). For clarity, cells are re-arranged in such a way that closer layers have smaller distance. (c) From top to bottom: intertwining compared to transport properties such as average return probability (rescaled by N), diffusion time and navigability, respectively, that can be enhanced on this multiplex network by using functional reducibility. Note that average return probability $\mathcal{R}(m)$ at reduction step m, integrated over τ, has been then transformed by $-\log(1 - \mathcal{R}(m)/\max_{m}\{\mathcal{R}(m)\})$ for clarity. (d) A multiplex network with 100 nodes where layers are uncorrelated random networks with wiring probability 0.2. (e) As in (b), but for the system considered in (d). (f) Transport properties are not enhanced by functional coupling because spectral diversity is maximized when the original system is not functionally aggregated. See Fig. 7 for results on additional synthetic models.

263 proteins characterizing the proteome of *Xenopus Laevis* (the African clawed frog) [37], where layers encode different genetic interactions (association, direct interaction, physical association, colocalization, and suppressive genetic interaction defined by inequality) [48]. Note that the last layer consists only of one link and we discard it from the analysis. Results are shown in Fig. 3(a) and summarized in Table I.

Our analysis suggests that, from a functional perspective, protein-protein interactions characterized by association and colocalization in this organism can be coupled together. This result is in striking disagreement with the result obtained from structural reducibility, where the other two layers were aggregated because indistinguishable from an information-theoretic perspective [37].

TABLE I. Functional coupling of layers in the multiplex protein-protein interaction network of *Xenopus Laevis*. Reducibility: $\chi_{\text{func}} = 33.3\%$.

Functional group	Layer ID (layer name)
1	1 (association)
	4 (colocalization)
2	2 (direct interaction)
3	3 (physical association)

C. Social system

The social system considered is the largest connected component of the collaboration network of the 475 scientists affiliated to the Pierre Auger experiment [46], the largest observatory of ultrahigh-energy cosmic rays. Here, collaborators work together in different research topics on specific tasks defining the 16 layers (e.g., source detection, mass composition, experimental enhancements, shower reconstruction, etc.) of the multilayer co-authorship system. Note that one layer was so sparse with respect to the other layers that we have discarded it from the analysis. Results are shown in Fig. 3(b) and summarized in Table II.

In this case, our analysis suggests the existence of two functional groups of layers. The two groups do not reflect similarities observed in the multilayer community organization of scientists [46], highlighting how the suggested functional coupling does not reflect trivial or complex structural features.

D. Urban and large-scale transportation systems

We further consider two transportation systems at different scales: the urban public transportation system of London consisting of 11 tube lines, DLR and Overground [6], and 369 stations, and the European airport network [47], consisting of 417 airports connected by flight routes served by 37 airlines
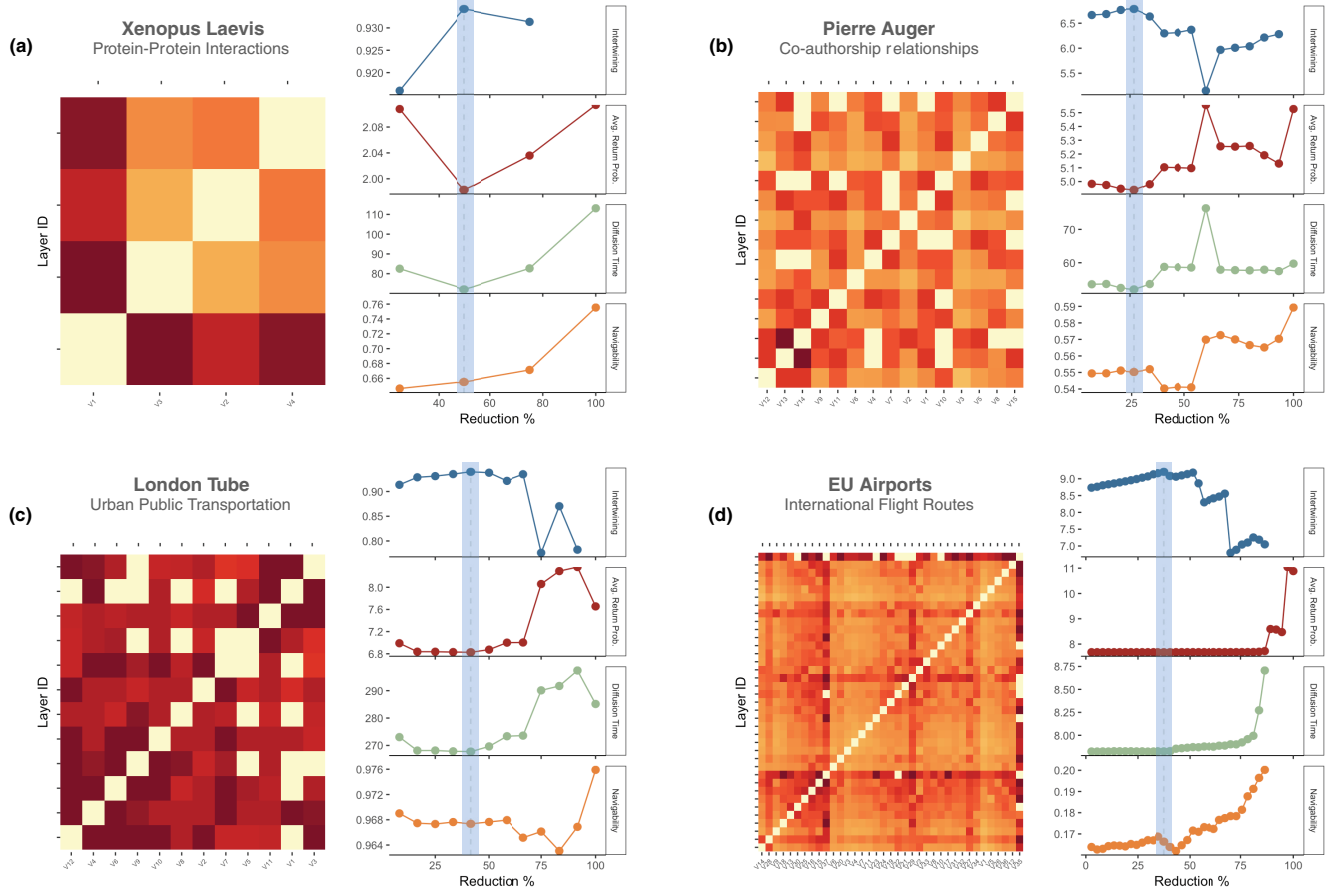
FIG. 3. Functional reducibility of empirical multilayer systems. The same analysis shown in Fig. 2 is performed on different multilayer systems, namely, the (a) protein-protein interactions of *Xenopus Laevis* (the African clawed frog) [37], (b) co-authorship network of the Pierre Auger experiment [46], (c) public transportation of London [6], and (d) European airport network [47]. See the text for further details about the data set. Note that entries d_{ij} of distance matrices have been transformed by $-\log(d_{ij})$ to enhance the visualization of the corresponding heat maps (The average return probability is rescaled by N.)

defining layers. Results are shown in Figs. 3(c) and 3(d) and summarized in Tables IV–III, respectively.

Remarkably, in the case of London transportation, the result differs from the one obtained from structural reducibility, which instead predicts the aggregation of only one layer

TABLE II. Similar to Table I, for the Pierre Auger collaboration network. Reducibility: $\chi_{\text{func}} = 21.4\%$.

Functional group	Layer ID (layer name)
1	5 (point source), 2 (detector), 9 (spectrum)
2	6 (mass composition), 12 (SD reconstruction)
3	1 (neutrinos)
4	3 (enhancements)
5	4 (anisotropy)
6	7 (horizontal)
7	8 (hybrid-reconstruction)
8	10 (photons)
9	11 (atmospheric)
10	13 (hadronic-interactions)
11	14 (exotics)
12	15 (magnetic)

TABLE III. Similar to Table I, for London Tube. Reducibility: $\chi_{\text{func}} = 33.3\%$.

Functional group	Layer ID (layer name)
1	2 (jubilee) 11 (circle)
2	4 (hammersmith and city) 10 (central)
3	6 (bakerloo) 12 (victoria)
4	7 (piccadilly) 8 (northern)
5	1 (dlr)
6	3 (district)
7	5 (overground)
8	9 (metropolitan)
9	13 (waterloo and city)

TABLE IV. Similar to Table I, for the EU airport network. Reducibility: $\chi_{\text{func}} = 35.1\%$.

Functional group	Layer ID (layer name)
1	2 (Ryanair)
	27 (Transavia Holland)
2	15 (Flybe)
	7 (Air France)
3	11 (Swiss International Air Lines)
	8 (Scandinavian Airlines)
4	20 (LOT Polish Airlines)
	10 (Alitalia)
5	23 (Air Lingus)
	12 (Iberia)
6	17 (TAP Portugal)
	13 (Norwegian Air Shuttle)
7	28 (Niki)
	16 (Wizz Air)
8	25 (Panagra Airways)
	18 (Brussels Airlines)
9	32 (European Air Transport)
	19 (Finnair)
10	33 (Malev Hungarian Airlines)
	21 (Vueling Airlines)
11	30 (Aegean Airlines)
	24 (Germanwings)
12	36 (TNT Airways)
	31 (Czech Airlines)
13	34 (Air Baltic)
14	1 (Lufthansa)
15	3 (Easyjet)
16	4 (British Airways)
17	5 (Turkish Airlines)
18	6 (Air Berlin)
19	9 (KLM)
20	14 (Austrian Airlines)
21	22 (Air Nostrum)
22	26 (Netjets)
23	29 (SunExpress)
24	35 (Wideroe)
	37 (Olympic Air)

($\chi_{\text{struct}} = 8\%$ [37]). This effect is due to the fact that distinguishability with respect to the aggregate network is not sufficient, alone, to capture the complex effects due to layer-layer coupling which, instead, are perfectly accounted for by the system's intertwining. In fact, the functional reducibility χ obtained in this case is 33.3%, with the advantage of lowering both the diffusion time and dynamical trapping. Interestingly, DLR and Overground, which are very different from other tube lines, are not functionally coupled with other layers, confirming that, from a functional perspective, they are not redundant for flowing passengers.

In the case of the EU airport network, we observe a functional reducibility of about 35%. It is worth mentioning that the analysis of structural reducibility of a larger data set—encoding flight routes in Europe from 175 airlines serving 1064 airports—reported a reduction of 6% [37]: we will discuss better this result in the next section.

It is worth discussing about results obtained for the navigability of all empirical systems considered here. In fact, this transport descriptor is not enhanced, at variance with the other descriptors, when intertwining is maximized. However, this is not a limitation of the proposed framework because the equation relating intertwining to navigability has been obtained under an approximation that, evidently, is not satisfied when many isolated state nodes are present. In fact, even the coverage, and consequently the navigability, has been originally introduced to deal with interconnected multilayer systems where state nodes exist in all layers. In this work, we have introduced its generalization to the case of noninterconnected multiplex networks: when state nodes are defined in all layers, as in our synthetic benchmarks, both the coverage and the navigability behave as expected and, in fact, are correctly enhanced. When a large fraction of state nodes are present only in a few layers, as in most empirical systems, random walkers are more unlikely to reach those nodes, dramatically reducing the overall probability of hitting the corresponding physical nodes. Since the study of a new measure of coverage and navigability is beyond the scope of this work, we leave it for future studies devoted to better define such measures to deal with more exotic topological scenarios.

V. CONCLUSIONS AND OUTLOOK

This study provides a transparent framework to better understand the interplay between complex topologies and noncompact search dynamics [49], allowing one to exploit the multifaceted interactions among constituents of complex systems—usually encoded by multilayer networks—to enhance transport phenomena without altering the underlying structure.

First, we defined the average dynamical trapping of a complex network to quantify the efficiency in information exchange between its constituents: the flow distribution is more efficient when this trapping is minimized. Therefore we demonstrated that this concept is intimately related to the spectral partition function of a complex network, allowing us to map the analysis of transport phenomena into the analysis of system's spectral entropy.

By applying the same principle to multilayer networks, we have identified a strategy for coupling subsets of layers—according to a functional criterion—in order to minimize the overall partition function. In fact, layers are grouped together by considering their superposition with respect to random walk dynamics. The choice of layers to superpose is driven by the deviation of system's entropy from the limit of noninteracting layers. This approach, justified from an information-theoretic perspective, has an elegant physical meaning: redundant diffusion pathways across layers are reduced until the multilayer system consists of layers with very different transport properties. This results is confirmed theoretically, showing that the average variance of the corresponding eigenvalue spectra must be maximum.

Results from synthetic benchmarks were in agreement with our expectations. Similarly, results from empirical multilayer systems confirmed the possibility to functionally couple group of layers to enhance information flow in the corresponding systems.

For instance, in the case of the collaboration network of Pierre Auger scientists, we provided evidence that some tasks should be coupled together. A mailing list—or any other shared communication mean—involving all scientists collaborating on the corresponding tasks would facilitate exchange of relevant information among peers. In fact, at the time described by the data set (i.e., between 2010 and 2012), an independent mailing list was assigned to each task and scientists had to subscribe to multiple ones to be updated with relevant information. However, this approach was not really efficient: it was not unusual to receive multiple times the same information.

The case of the European airport network was interesting as well. From a policy perspective, the functional coupling of routes corresponding to different airlines suggests that passenger flows would benefit from shared tickets. Among the identified functional groups, many are very reasonable when the location of the corresponding headquarters is considered: e.g., Air France and Swiss International Air Lines, Iberia and TAP Portugal, Ryanair and Flybe. Remarkably, while airline companies tend to avoid overlapping routes, as demonstrated by a structural reducibility of 6% [37], they are correlated (functional reducibility of 35%) in a such a way that, from a functional point of view, they can still improve their transportation service.

In the case of London public transport, we have found that a few configurations close to the functionally optimal one are also plausible in terms of enhanced transport properties and intertwining. This result was not surprising: the layers of this multiplex system have treelike structures which tend to avoid topological overlap towards peripheral areas, thus different functional configurations might enhance transport in a similar way.

On the one hand, our findings open the doors to a broad spectrum of applications where superimposing layers with tailored policies induces relevant and controllable changes in transport phenomena, reducing diffusion time and enhancing navigability in multilayer networks. Nevertheless, the proposed approach can be easily adapted to scenarios where one physically acts on system's topology, not necessarily multilayer, for instance by adding (or removing) either its constituents or its connections. On the other hand, the theoretical framework developed in this work can be used further explore the bridge between statistical physics and network information theory.

APPENDIX A: BASIC NOTATION

Given a complex network of N nodes, undirected and weighted, the corresponding normalized Laplacian matrix governing the random walk dynamics, is defined by

$$\mathcal{L} = I - D^{-1}W, \tag{A1}$$

W being the adjacency matrix of the network and D the diagonal matrix with entries $D_{kk} = s_k = \sum_{i=1}^{N} W_{ki}$.

The corresponding density matrix [35] is defined by

$$\rho(\tau) = \frac{e^{-\tau\mathcal{L}}}{\mathrm{Tr}(e^{-\tau\mathcal{L}})}. \tag{A2}$$

The spectral entropy of the network is therefore calculated according to the definition of Von Neumann entropy in case of a quantum system:

$$
\begin{aligned}
S(\tau) &= -\mathrm{Tr}(\rho(\tau)\log_2\rho(\tau)) \\
&= \frac{1}{Z(\tau)\log 2}\mathrm{Tr}(\tau\mathcal{L}e^{-\tau\mathcal{L}}) + \log_2 Z(\tau),
\end{aligned}
$$

where $Z(\tau) = \mathrm{Tr}(e^{-\tau\mathcal{L}})$ is the partition function.

In the following, we will use the notation $Z_{\mathrm{int}}(\tau)$ and $S_{\mathrm{int}}(\tau)$ to indicate, respectively, partition function and spectral entropy of a multiplex network. We will avoid to specify the index where this choice does not generate ambiguity.

If the multiplex consists of L layers, we will use the notation $Z^{(\ell)}(\tau)$ and $S^{(\ell)}(\tau)$ for each layer $\ell = 1, 2, \ldots, L$ separately.

APPENDIX B: RANDOM WALK DYNAMICS ON MULTIPLEX

Random walk dynamics on top of multiplex systems has not been yet defined. Therefore, for our purposes, we firstly introduce a more general framework to derive the multiplex master equation as a function of diffusion time τ.

In fact, when exploring a layer ℓ in isolation, a random walker jumps from node i to node j with probability $T_{ij}^{(\ell)} = \frac{A_{ij}^{(\ell)}}{k_i^{(\ell)}}$, where $\mathbf{T}^{(\ell)}$ is the transition matrix of the random walk, $\mathbf{A}^{(\ell)}$ is the adjacency matrix of the layer and $k_i^{(\ell)} = \sum_{j=1}^{N} A_{ij}^{(\ell)}$ is degree of node i in layer ℓ. Interacting layers in a multiplex network exhibit a richer dynamics, as the random walker in node i can first switch from layer ℓ' to layer ℓ with probability $P_{\ell'\ell}^{(i)}$, and then jump from node i to node j with probability $T_{ij}^{(\ell)}$. Although all results of this paper are valid for any probability distribution, in absence of knowledge about the relative importance of weights, let us assume in the following that the probability to switch layer is uniform, i.e., $P_{\ell'\ell} = 1/L$: then the overall probability to observe a transition from i to j, regardless of the layer, is given by $T_{ij} = \frac{1}{L}\sum_{\ell=1}^{L} T_{ij}^{(\ell)}$, which can be written in a more compact notation as $\mathbf{T} = \langle\mathbf{T}^{(\ell)}\rangle$. The weighted average, in case of further knowledge about the layer's weights, becomes: $\langle\mathbf{T}^{(\ell)}\rangle = \sum_{\ell=1}^{L} P^{(\ell)}T^{(\ell)}$, where $P^{(\ell)}$ is the contribution of layer ℓ in the total flow.

It is worth remarking once again that the dynamics is dramatically different from aggregating all layers to a single network and then calculating the transition matrix of the random walk. In fact, two layers encode networks which are represented by adjacency matrices: aggregating two layers is achieved by summing entrywise their adjacency matrices. Similarly, superimposing layers makes them indistinguishable to the eyes of random walker, so the dynamics can be seen as random walk on the aggregated network. Evidently, the distinguishability of layers' dynamics, plays an important role in transport phenomena of complex systems and cannot be simply neglected. In the following, we will use the terms aggregation and superposition interchangeably. Two or more layers are functionally grouped together, or aggregated, if the corresponding networks are superimposed.

Having the transition matrix, we can find the Laplacian matrix of multiplex given by $\mathcal{L} = \mathbf{I} - \mathbf{T}$. This allows for finding the master equation for random walk on multiplex networks. If the ith components of the vector $\mathbf{p}(\tau)$ indicates the probability to find the random walker in node i at time τ, its evolution is governed by the master equation

$$\mathbf{p}(\tau + 1) = \mathbf{p}(\tau)\mathbf{T}, \quad (B1)$$

which in the continuous-time approximation, reduces to [Eq. (B2)]

$$\frac{\partial \mathbf{p}(\tau)}{\partial \tau} + \mathbf{p}(\tau)\mathcal{L} = 0, \quad (B2)$$

with solution given by $\mathbf{p}(\tau) = \mathbf{p}(0)e^{-\tau\mathcal{L}}$. Here, $\mathcal{L}$ indicates the normalized Laplacian matrix of the multiplex, $\mathbf{p}(\tau)$ is the vector encoding the probability of finding the random walker in a specific node at time τ, and $e^{-\tau\mathcal{L}}$ is the propagator of random walk dynamics. Note that the Laplacian matrix of the multiplex network is derived in terms of the average of Laplacian matrices $\mathcal{L}^{(\ell)}$ ($\ell = 1, 2, \ldots, L$) of the single layers, i.e., $\mathcal{L} = \langle \mathcal{L}^{(\ell)} \rangle$.

APPENDIX C: DYNAMICAL TRAPPING

To show this, let $\mathbf{p}(0) = \mathbf{p}(i|0) \equiv (0, \ldots, 1, 0, \ldots, 0)$ indicate that the walk originates in node i at time 0 with probability 1. In practice $\mathbf{p}(i|0) = \mathbf{e}_i$ is the ith canonical vector in $\mathbb{R}^N$, being N the size of the network. The *return probability* for node i at time τ is the probability of finding the random walker in i at time τ steps later assuming it originated in i at time 0, i.e., $\mathbf{p}(\tau)\mathbf{p}^\dagger(i|0)$, where † indicates the transpose operator. From the solution of Eq. (B2) in terms of the random walk propagator:

$$\mathbf{p}(\tau)\mathbf{p}^\dagger(i|0) = \mathbf{p}(0)e^{-\tau\mathcal{L}}\mathbf{p}^\dagger(0) = (e^{-\tau\mathcal{L}})_{ii}. \quad (C1)$$

Consequently, the *average return probability* is given by

$$\mathcal{R}(\tau) = \frac{1}{N}\sum_{i=1}^{N}(e^{-\tau\mathcal{L}})_{ii} = \frac{1}{N}\mathrm{Tr}(e^{-\tau\mathcal{L}}) = \frac{Z(\tau)}{N}. \quad (C2)$$

APPENDIX D: PROPERTIES OF THE PARTITION FUNCTION

Spectral entropy can be used to compare two networks [35]. One measure of similarity is the Kullback-Leibler (KL) divergence. Spectral KL divergence is non-negative, exactly as its information-theoretic counterpart.

Given a multiplex network with density matrix $\boldsymbol{\rho}(\tau)$ and any of its layers, indicated by $l = 1, 2, \ldots, L$, the following inequality holds:

$$\mathcal{D}_{KL}(\boldsymbol{\rho}||\boldsymbol{\rho}^{(\ell)}) = \mathrm{Tr}(\boldsymbol{\rho}(\log_2 \boldsymbol{\rho} - \log_2 \boldsymbol{\rho}^{(\ell)})) \geqslant 0. \quad (D1)$$

Note that the explicit dependence on τ is omitted for sake of clarity.

The average entropy distance of multiplex and its layers, can be quantified by the Kullbeck-Leibler entropy divergence, and in case of uniform distribution of layer weights, defines

the intertwining as follows:

$$\mathcal{I} = \langle \mathcal{D}_{KL}(\boldsymbol{\rho}||\boldsymbol{\rho}^{(\ell)}) \rangle = \frac{1}{L}\sum_{\ell=1}^{L}\mathcal{D}_{KL}(\boldsymbol{\rho}||\boldsymbol{\rho}^{(\ell)})$$

$$= \frac{1}{L}\sum_{\ell=1}^{L}\log(Z^{(\ell)}) - \log Z, \quad (D2)$$

which is zero if the system as a whole does not provide any additional information about its layers in isolation. Conversely, a large average divergence between the system and its layers highlights the necessity for using the multiplex description to encode the layer-layer interactions. In the next sections, we will show how the normalized version of this measure can be understood in terms of the entropy of layers in isolation and multiplex as a whole, and how it can be used to enhance transport properties while superimposing subsets of layers. We emphasize again that also intertwining can be found for any arbitrary set of layer weights, in case we have a way to asses the relative importance of layers.

Using Eq. (A3), we obtain

$$\tau\mathrm{Tr}(\mathcal{L}^{(\ell)}\boldsymbol{\rho}) + \log Z^{(\ell)} - \tau\mathrm{Tr}(\mathcal{L}\boldsymbol{\rho}) - \log Z \geqslant 0. \quad (D3)$$

As it is true for all layers, we sum over layers and divide by the number of layers and exploit the relation $\mathcal{L} = \frac{1}{L}\sum_{\ell=1}^{L}\mathcal{L}^{(\ell)}$ to obtain the average divergence

$$\frac{1}{L}\sum_{\ell=1}^{L}\mathcal{D}_{KL}(\boldsymbol{\rho}||\boldsymbol{\rho}^{(\ell)}) = \log\left(\prod_{\ell=1}^{L}Z^{(\ell)}\right)^{\frac{1}{L}} - \log Z \geqslant 0,$$

from which the following fundamental inequality, relating the partition functions of layers with the one of the multiplex system, is obtained

$$Z(\tau) \leqslant \prod_{\ell=1}^{L}Z^{(\ell)}(\tau)^{\frac{1}{L}}. \quad (D4)$$

A direct consequence of this result is that at least one layer $\ell^\star$ must have a partition function such that $Z^{(\ell^\star)}(\tau) \geqslant Z(\tau)$, or inequality in Eq. (D4) would not be satisfied. It follows:

$$\log Z^{(\ell^\star)}(\tau) \geqslant \log Z(\tau), \quad (D5)$$

and

$$\tau\frac{Z^{(\ell^\star)}(\tau) - 1}{Z^{(\ell^\star)}(\tau)} \geqslant \tau\frac{Z(\tau) - 1}{Z(\tau)}, \quad (D6)$$

which will be useful in the following.

This last result holds for networks consisting of a single connected component, a scenario that could not be satisfied by many empirical systems. Its generalization to the case of multiple connected components is obtained as follows. First, let us assume that there are C connected components in the system. If $\tau \gg 1$ then $Z(\tau) \approx C$ and the following approximation holds:

$$\log Z(\tau) = -\log\frac{1}{Z(\tau)} \approx \log C + \frac{Z(\tau) - C}{Z(\tau)}, \quad (D7)$$

leading to

$$\tau\log C^{(\ell^\star)} + \tau\frac{Z^{(\ell^\star)}(\tau) - C^{(\ell^\star)}}{Z^{(\ell^\star)}(\tau)} \geqslant \tau\log C + \tau\frac{Z(\tau) - C}{Z(\tau)}.$$

APPENDIX E: MEAN-FIELD APPROXIMATION

The eigenvalue spectrum of the normalized Laplacian satisfies the following properties: (1) $0 = \lambda_1 \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_N \leqslant 2$ and (2) $\mathrm{Tr}(\mathcal{L}) = \sum_{i=1}^{N} \lambda_i = N$. At this step, it is worth noting that ρ and $\mathcal{L}$ can be eigendecomposed as follows:

$$\mathcal{L} = Q \Lambda Q^{-1}, \qquad (E1)$$

$$\rho(\tau) = Q \frac{e^{-\tau \Lambda}}{Z(\tau)} Q^{-1}, \qquad (E2)$$

being the columns of Q the eigenvectors of the normalized Laplacian matrix and Λ a diagonal matrix with $\Lambda_{ii} = \lambda_i$. It follows that $\nu_i(\tau) = e^{-\tau \lambda_i}/Z(\tau)$ and, as a consequence,

$$\mathrm{Tr}(\mathcal{L}\rho) = \sum_{i=1}^{N} \lambda_i \nu_i(\tau) = \sum_{i=C+1}^{N} \lambda_i \frac{e^{-\tau \lambda_i}}{Z(\tau)}, \qquad (E3)$$

the last step justified by the fact that $\lambda_1, \ldots, \lambda_C = 0$ for a network with C connected components, which includes the possibility for C_{isol} isolated nodes, i.e., nodes not interconnected with the rest of the system. For sake of completeness, it is worth remarking that $\mathrm{Tr}(\mathcal{L}) = N$ if there are multiple connected components but no isolated nodes, whereas $\mathrm{Tr}(\mathcal{L}) = N - C_{\mathrm{isol}}$ if isolated nodes are present.

The mean-field approximation consists in neglecting higher-order terms in the following:

$$\begin{aligned} \langle \lambda \nu(\tau) \rangle &= \langle (\lambda - \bar{\lambda} + \bar{\lambda})(\nu(\tau) - \bar{\nu}(\tau) + \bar{\nu}(\tau)) \rangle \\ &= \bar{\lambda}\bar{\nu}(\tau) + \langle (\lambda - \bar{\lambda})(\nu(\tau) - \bar{\nu}(\tau)) \rangle \\ &\approx \bar{\lambda}\bar{\nu}(\tau). \end{aligned} \qquad (E4)$$

To calculate the mean values, it is worth noting that

$$\bar{\lambda} = \frac{1}{N-C} \sum_{i=C+1}^{N} \lambda_i = \frac{N - C_{\mathrm{isol}}}{N - C} \qquad (E5)$$

and

$$\bar{\nu}(\tau) = \frac{1}{N-C} \sum_{i=C+1}^{N} \frac{e^{-\tau \lambda_i}}{Z(\tau)} = \frac{1}{N-C} \frac{Z(\tau) - C}{Z(\tau)}. $$

It follows that

$$\mathrm{Tr}(\mathcal{L}\rho) = (N-C)\langle \lambda \nu(\tau) \rangle \approx \frac{N - C_{\mathrm{isol}}}{N - C} \frac{Z(\tau) - C}{Z(\tau)},$$

which, for a network with no isolated nodes and only one connected component, it reduces to

$$\mathrm{Tr}(\mathcal{L}\rho) \approx \frac{Z(\tau) - 1}{Z(\tau)}, \qquad (E6)$$

for $N \gg 1$. It follows that in the thermodynamic limit the mean-field entropy is given by

$$S_{\mathrm{MF}}^{\mathrm{therm}}(\tau) = \frac{1}{\log 2} \left(\tau \frac{Z(\tau) - 1}{Z(\tau)} + \log Z(\tau) \right), \qquad (E7)$$

or, more generally,

$$S_{\mathrm{MF}}(\tau) = \frac{1}{\log 2} \left(\tau \frac{N - C_{\mathrm{isol}}}{N - C} \frac{Z(\tau) - C}{Z(\tau)} + \log Z(\tau) \right),$$
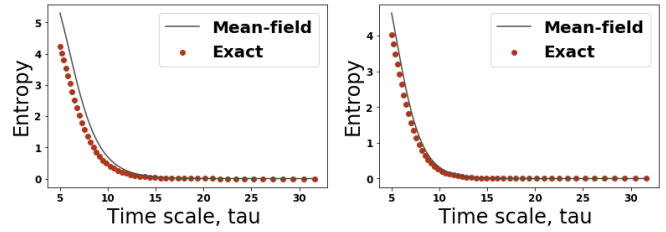

FIG. 4. Entropy of Barabasi-Albert networks with $m = 5$ (left) and 8 (right) compared with the mean-field expectation.

which, in the thermodynamic limit, reduces to

$$S_{\mathrm{MF}}^{\mathrm{therm}}(\tau) = \frac{1}{\log 2} \left(\tau \frac{Z(\tau) - C}{Z(\tau)} + \log Z(\tau) \right), \qquad (E8)$$

if C scales sublinearly with N, i.e., if the number of isolated nodes and disconnected components is much smaller than the size of the system, which might not be the case for finite empirical systems.

Taylor expanding the logarithm, keeping the first term in the limit of large τ,

$$\begin{aligned} S_{\mathrm{MF}}^{\mathrm{therm}}(\tau) &= \frac{1}{\log 2} \left(\tau \frac{Z(\tau) - C}{Z(\tau)} + \log C + \frac{Z(\tau) - C}{Z(\tau)} \right) \\ &= \frac{1}{\log 2} \left((\tau + 1) \frac{Z(\tau) - C}{Z(\tau)} + \log C \right) \\ &= \frac{(\tau + 1)}{\log 2} \frac{Z(\tau) - C}{Z(\tau)} + \log_2 C, \qquad (E9) \end{aligned}$$

The goodness of the mean-field approximation is numerically investigated for a variety of network models (see Figs. 4 and 5 for emblematic examples).

APPENDIX F: FIRST-ORDER CORRECTION

In the previous section, we have neglected the cross-correlation term $\langle (\lambda - \bar{\lambda})(\nu(\tau) - \bar{\nu}(\tau)) \rangle$ to obtain the mean-field approximation of the spectral entropy. It is straightforward to show that this term can be considered as a correction to the mean field:

$$\begin{aligned} \delta S_{\mathrm{MF}}(\tau) &= \frac{\tau}{\log 2} \langle (\lambda - \bar{\lambda})(\nu(\tau) - \bar{\nu}(\tau)) \rangle \\ &= \frac{1}{\log 2} \frac{\tau e^{-\tau}}{Z(\tau)} \sum_{i=2}^{N} \delta_i e^{-\delta_i \tau}, \qquad (F1) \end{aligned}$$
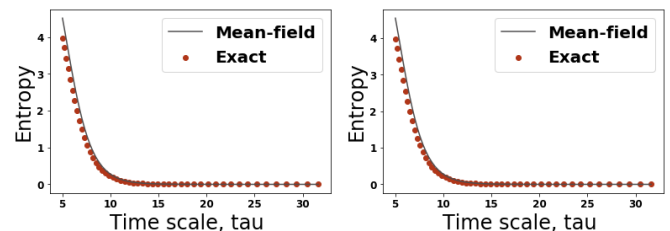

FIG. 5. Entropy of Erdos-Renyi networks with $p = 0.2$ (left) and 0.5 (right) compared with the mean-field expectation.

in order to write the spectral entropy as $S(\tau) = S_{\mathrm{MF}}(\tau) + \delta S_{\mathrm{MF}}(\tau)$. Note that $\delta_i = \lambda_i - \langle\lambda\rangle$, where λ_i is the i-th eigenvalue of the normalized Laplacian matrix and $\langle\lambda\rangle = 1$ in the thermodynamic limit.

In fact, $Z(\tau) \geqslant 1$ for any value of τ, therefore the following inequality holds:

$$\delta S_{\mathrm{MF}}(\tau) \leqslant -\tau^2 e^{-\tau}(N-1)(\langle\lambda^2\rangle - 1), \qquad (\mathrm{F2})$$

where we have expanded the exponential in its Taylor series truncated at second order and we have used the fact that $\langle\delta^2\rangle = \langle\lambda^2\rangle + 1$. In the case of C disconnected components, this difference is

$$\delta S_{\mathrm{MF}}(\tau) \leqslant -\tau^2 e^{-\tau}(N-C)(\langle\lambda^2\rangle - 1). \qquad (\mathrm{F3})$$

From the definition of $\bar{S}_{\mathrm{nint}}(\tau)$ and the above results, it is straightforward to show that

$$\bar{S}_{\mathrm{nint}}(\tau) \leqslant \frac{1}{L}\sum_{\ell=1}^{L} S_{\mathrm{MF}}^{(\ell)}(\tau) - \tau^2 e^{-\tau}\frac{N-1}{L}\sum_{\ell=1}^{L}\big(\langle\lambda^{(\ell)^2}\rangle - 1\big),$$

$$S_{\mathrm{int}}(\tau) \leqslant S_{\mathrm{int,MF}}(\tau) - \tau^2 e^{-\tau}(N-1)\big(\langle\lambda_{\mathrm{int}}^2\rangle - 1\big).$$

For a stochastic variable X bounded in an interval $[x_m, x_M]$ the Popoviciu's inequality holds:

$$\mathrm{Var}(X) \leqslant \frac{(x_M - x_m)^2}{4}, \qquad (\mathrm{F4})$$

that in our case reduces to $\mathrm{Var}(\lambda) \leqslant 1$, for the distribution of normalized Laplacian matrix's eigenvalues. It follows that $0 \leqslant \langle\lambda^2\rangle = \mathrm{Var}(\lambda) + \langle\lambda\rangle^2 \leqslant 2$. For $\tau \gg 1$, then

$$\bar{S}_{\mathrm{nint}}(\tau) \leqslant \frac{1}{L}\sum_{\ell=1}^{L} S_{\mathrm{MF}}^{(\ell)}(\tau) \qquad (\mathrm{F5})$$

$$S_{\mathrm{int}}(\tau) \leqslant S_{\mathrm{int,MF}}(\tau). \qquad (\mathrm{F6})$$

APPENDIX G: NON-NEGATIVITY OF INTERTWINING

From the fundamental inequality, it's straightforward to show that

$$\log Z_{\mathrm{int}}(\tau) \leqslant \frac{1}{L}\sum_{\ell=1}^{L}\log Z^{(\ell)}(\tau). \qquad (\mathrm{G1})$$

If $\tau \gg 1$ then $Z(\tau) \approx 1$ and the following approximation holds:

$$\log Z(\tau) = -\log\frac{1}{Z(\tau)} \approx \frac{Z(\tau) - 1}{Z(\tau)}, \qquad (\mathrm{G2})$$

which leads to

$$\frac{Z_{\mathrm{int}}(\tau) - 1}{Z_{\mathrm{int}}(\tau)} \leqslant \frac{1}{L}\sum_{\ell=1}^{L}\frac{Z^{(\ell)}(\tau) - 1}{Z^{(\ell)}(\tau)}. \qquad (\mathrm{G3})$$

By summing side by side inequalities (G1) and (G3), we obtain

$$S_{\mathrm{int}}(\tau) \leqslant \bar{S}_{\mathrm{nint}}(\tau), \qquad (\mathrm{G4})$$

leading to non-negativity of the intertwining

$$\mathcal{I}(\tau) = \bar{S}_{\mathrm{nint}}(\tau) - S_{\mathrm{int}}(\tau) \geqslant 0, \qquad (\mathrm{G5})$$

for $\tau \gg 1$. We wonder to which extent this inequality holds for smaller values of τ.

After expanding Z around $\delta_i = \lambda_i - \langle\lambda\rangle = \lambda_i - 1$, we have at first order

$$\begin{aligned}
Z(\tau) &= 1 + \sum_{i=2}^{N} e^{-\tau\lambda_i} = 1 + e^{-\tau}\sum_{i=2}^{N} e^{-\tau\delta_i}\\
&= 1 + (N-1)e^{-\tau}\left\langle 1 - \tau\delta + \frac{\tau^2}{2}\delta^2 - \dots \right\rangle\\
&\approx 1 + (N-1)e^{-\tau}, \qquad (\mathrm{G6})
\end{aligned}$$

from which $Z^{-1}(\tau) \approx 1 - (N-1)e^{-\tau}$.

It follows that Eq. (G2) is satisfied provided that $(N-1)e^{-\tau} \ll 1$, which is equivalent to require that τ should be larger than the τ_c where $\tau_c > \log(N-1)$ in order for mean-field approximation to be very accurate.

In the more general case, we have $Z(\tau) \approx C + (N-C)e^{-\tau}$, from which

$$Z^{-1}(\tau) \approx C^{-1}\left(1 - \frac{N-C}{C}e^{-\tau}\right), \qquad (\mathrm{G7})$$

which similarly holds if $\tau_c > \log(\frac{N}{C} - 1)$ so that $\log\frac{Z}{C}$ can be expanded in the same way. Note that if there are isolated nodes, the fundamental inequality must be proven.

APPENDIX H: WHEN AND WHY THE STRUCTURAL REDUCIBILITY FAILS

Of course, the noninteracting scenario described here is barely observed in reality and $S_{\mathrm{nint}}(\tau)$ would be equal to the entropy of one single layer. However, our framework suggests that even in the case of different layers, we can interpret $S_{\mathrm{nint}}(\tau)$ as an upper bound—obtained by exploring the system using each layer separately—to $S_{\mathrm{int}}(\tau)$ (see Appendices E–G). Then, any deviation from this upper bound highlights the existence of layer-layer interactions and can be used functional reduciblity of the system. In the light of this distinction, we are now able to better understand existing measures in the literature. For instance, the classical structural reducibility [37] is based on comparing S_{nint} against the entropy of the fully aggregated system S_{agg}: clearly this method has to fail when some layers are exactly the same and, more generally, when they are very similar with each other (Appendix H).

The classical structural reducibility method is based on maximizing the quality function

$$q(n) = 1 - \frac{\langle S^{(\ell)}\rangle(n)}{S_{\mathrm{agg}}}, \qquad (\mathrm{H1})$$

where n is the aggregation step, and S_{agg} is the entropy of the fully aggregated system. Let us consider a multiplex network with L layers, where m layers are identical and have entropy $S^{(\mathrm{same})} = S^{(\ell)}$ ($\ell = 1, 2, \dots, m$), while the remaining layers have different topologies and we indicate by

$$S^{(\mathrm{dif})} = \frac{1}{L-m}\sum_{\ell=L-m+1}^{L} S^{(\ell)} \qquad (\mathrm{H2})$$

their average entropy. We expect a suitable optimization procedure to identify the $m - 1$ redundant layers and reduce the multiplex system to $L - m + 1$ layers.

During the first m aggregation steps, the layers that are supposed to be aggregated are exactly the same. Remarking that the entropy remains unchanged if the adjacency matrix is multiplied by a constant, when two out of m equal layers are aggregated, the entropy of the resulting network is still $S^{(\text{same})}$. So when calculating the average entropy of layers, aggregation of every pair of equal layers is exactly equal to removing one of them. This way it is straightforward to write the average entropy of layers as a function of aggregation step n which follows:

$$
\begin{aligned}
\langle S^{(\ell)} \rangle (n) &= \frac{1}{L-n} \sum_{\ell=n+1}^{m} S^{(\ell)} \\
&= \frac{1}{L-n} \left[\sum_{\ell=n+1}^{m} S^{(\ell)} + \sum_{\ell=L-m+1}^{L} S^{(\ell)} \right] \\
&= \frac{m-n}{L-n} S^{(\text{same})} + \frac{L-m}{L-n} S^{(\text{dif})} \\
&= \frac{1}{L-n} \left[(m-n) S^{(\text{same})} + (L-m) S^{(\text{dif})} \right].
\end{aligned}
$$

Writing one of the entropies in terms of the other one $S^{(\text{dif})} = \delta + S^{(\text{same})}$, we have

$$
\begin{aligned}
\langle S^{(\ell)} \rangle (n) &= \frac{1}{L-n} \big[(m-n) S^{(\text{same})} \\
&\quad + (L-m) S^{(\text{same})} + (L-m)\delta \big] \\
&= S^{(\text{same})} + \frac{L-m}{L-n} \delta.
\end{aligned}
\tag{H3}
$$

Therefore, if $S^{(\text{dif})} > S^{(\text{same})}$ then δ is positive, and the aggregation of similar layers increases the average entropy of layers $\langle S^{(\ell)} \rangle (n)$, while decreasing the classical quality function: $q(n) = 1 - \frac{\langle S^{(\ell)} \rangle (n)}{S_{\text{agg}}}$. This case corresponds to an irreducible structure, an evident failure of the method because it is only comparing the average entropy of different groups of layers. A convincing example is given in Fig. 6, where the method is applied to a multiplex system consisting of ten layers (nine are identical).

APPENDIX I: USING INTERTWINING FOR DIMENSIONALITY REDUCTION

If $\mathcal{L}$ indicates the normalized Laplacian matrix of the multiplex network, let us introduce perturbation matrices $\Delta \mathcal{L}^{(\ell)}$ such that the normalized Laplacian matrix of each layer can be written as

$$
\mathcal{L}^{(\ell)} = \mathcal{L} + \Delta \mathcal{L}^{(\ell)}.
\tag{I1}
$$

What is the effect of such perturbations on the corresponding partition functions? In fact, we have already shown that

$$
Z(\tau) \simeq 1 + (N-1)e^{-\tau} \left[1 + \frac{\tau^2}{2} (\langle \lambda^2 \rangle - 1) - \dots \right]
\tag{I2}
$$

either for the multiplex network or for each layer separately. This result suggests that it is sufficient to find a relation
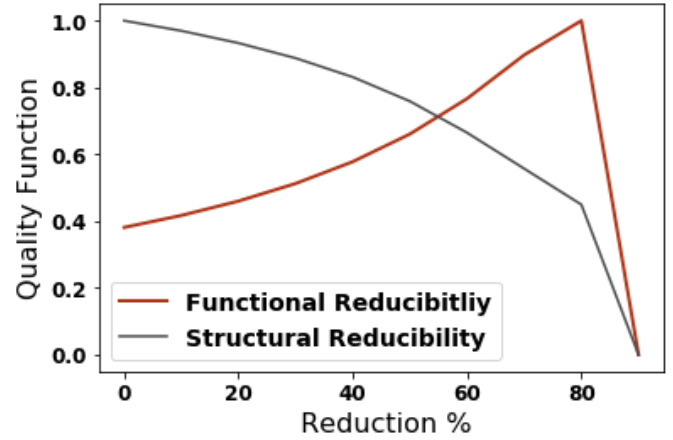

FIG. 6. Functional vs structural redundancy. A multiplex network with ten layers is given, in which nine layers are exactly the same and expected to be grouped into one network. In scenarios like this one, the structural reducibility fails to identify the redundant layers to superimpose. The curve corresponding to functional redundancy has been rescaled to match the scale of the other curve to facilitate the comparison of trends and the identification of maxima.

between the second-order moments of the eigenvalue distribution for networks separately and the whole multiple system.

In the following, we use the perturbation approach suggested by Lord Rayleigh, where we implicitly assume that perturbations to the Laplacian matrices are small. Since matrices $\mathcal{L}^{(\ell)}$ are diagonalizable, then

$$
\lambda_i^{(\ell)} = \lambda_i + \Delta \lambda_i^{(\ell)},
\tag{I3}
$$

with

$$
\Delta \lambda_i^{(\ell)} = \mathbf{q}_i^\top \Delta \mathcal{L}^{(\ell)} \mathbf{q}_i,
\tag{I4}
$$

where $\mathbf{q}_i$ indicates the ith eigenvector of $\mathcal{L}$. Since $\langle \lambda^{(\ell)} \rangle = 1$ and $\langle \lambda \rangle = 1$, it follows that $\langle \Delta \lambda^{(\ell)} \rangle = 0$, whereas

$$
\langle (\lambda^{(\ell)})^2 \rangle = \langle \lambda^2 \rangle + \langle (\Delta \lambda^{(\ell)})^2 \rangle,
\tag{I5}
$$

where the covariance term $\langle \lambda \Delta \lambda^{(\ell)} \rangle = 0$ because perturbed eigenvalues are not depending from the eigenvalues of $\mathcal{L}$. It follows that

$$
Z^{(\ell)}(\tau) = Z_{\text{int}}(\tau) + \Delta Z^{(\ell)}(\tau),
\tag{I6}
$$

with

$$
\Delta Z^{(\ell)}(\tau) = \frac{\tau^2}{2} e^{-\tau} (N-1) \langle (\Delta \lambda^{(\ell)})^2 \rangle.
\tag{I7}
$$

We can also write the product of partition functions of layers in terms of perturbations:

$$
\prod_{\ell=1}^{L} Z^{(\ell) \frac{1}{L}}(\tau) = \prod_{\ell=1}^{L} \left(Z_{\text{int}}(\tau) + \Delta Z^{(\ell)}(\tau) \right)^{\frac{1}{L}}
\tag{I8}
$$

$$
= Z_{\text{int}}(\tau) \prod_{\ell=1}^{L} \left(1 + \frac{\Delta Z^{(\ell)}(\tau)}{Z_{\text{int}}(\tau)} \right)^{\frac{1}{L}}.
\tag{I9}
$$

Since $1 \gg \frac{\Delta Z^{(\ell)}(\tau)}{Z_{\text{int}}(\tau)}$ we can use the approximation $(1 + \frac{\Delta Z^{(\ell)}(\tau)}{Z_{\text{int}}(\tau)})^{\frac{1}{L}} \approx 1 + \frac{1}{L}\frac{\Delta Z^{(\ell)}(\tau)}{Z_{\text{int}}(\tau)}$ at first order, to obtain

$$
\prod_{\ell=1}^{L}\left(1 + \frac{\Delta Z^{(\ell)}(\tau)}{Z_{\text{int}}(\tau)}\right)^{\frac{1}{L}} \approx \prod_{\ell=1}^{L}\left(1 + \frac{1}{L}\frac{\Delta Z^{(\ell)}(\tau)}{Z_{\text{int}}(\tau)}\right)
$$

$$
= 1 + \frac{1}{L}\sum_{\ell=1}^{L}\frac{\Delta Z^{(\ell)}(\tau)}{Z_{\text{int}}(\tau)}
$$

$$
+ \mathcal{O}\left(\Delta Z^{(\ell)^2}(\tau), \dots\right)
$$

$$
\approx 1 + \overline{\Delta Z^{(\ell)}}(\tau), \tag{I10}
$$

where

$$
\overline{\Delta Z^{(\ell)}}(\tau) \approx \frac{\tau^2}{2}e^{-\tau}(N-1)\overline{\langle(\Delta\lambda^{(\ell)})^2\rangle}
$$

and

$$
\overline{\langle(\Delta\lambda^{(\ell)})^2\rangle} = \frac{1}{L}\sum_{\ell=1}^{L}\langle(\Delta\lambda^{(\ell)})^2\rangle.
$$

Above the time τ_c, Eq. (G2) holds and we can write entropy in mean-field regime as

$$
S_{\text{MF}}(\tau) \times \log 2 = \tau\frac{Z(\tau)-1}{Z(\tau)} + \log Z(\tau)
$$

$$
\approx (\tau+1)\frac{Z(\tau)-1}{Z(\tau)}. \tag{I11}
$$

On the one hand, let us consider the average information divergence given by

$$
\frac{1}{L}\sum_{\ell=1}^{L}\mathcal{D}_{KL}(\boldsymbol{\rho}||\boldsymbol{\rho}^{(\ell)}) = \frac{1}{L}\sum_{\ell=1}^{L}\log\left(Z^{(\ell)}\right) - \log Z,
$$

where, multiplying by $\tau + 1$ we obtain the important relation with the entropy deviation given by

$$
(\tau+1)\frac{1}{L}\sum_{\ell=1}^{L}\mathcal{D}_{KL}(\boldsymbol{\rho}||\boldsymbol{\rho}^{(\ell)}) = (\tau+1)\frac{1}{L}\sum_{\ell=1}^{L}\log\left(Z^{(\ell)}\right)
$$

$$
- (\tau+1)\log Z
$$

$$
= \bar{S}_{\text{nint}}(\tau) - S_{\text{int}}(\tau).
$$

On the other hand, since

$$
\log\frac{\prod_{\ell=1}^{L}Z^{(\ell)\frac{1}{L}}(\tau)}{Z_{\text{int}}(\tau)} = \log\left(1 + \frac{\overline{\Delta Z^{(\ell)}}(\tau)}{Z_{\text{int}}(\tau)}\right) \approx \frac{\overline{\Delta Z^{(\ell)}}(\tau)}{Z_{\text{int}}(\tau)},
$$

we can write

$$
(\tau+1)\log\frac{\prod_{\ell=1}^{L}Z^{(\ell)}(\tau)^{\frac{1}{L}}}{Z_{\text{int}}(\tau)} = (\tau+1)\frac{1}{L}\sum_{\ell=1}^{L}\frac{Z^{(\ell)}(\tau)-1}{Z^{(\ell)}(\tau)}
$$

$$
- (\tau+1)\frac{Z_{\text{int}}(\tau)-1}{Z_{\text{int}}(\tau)},
$$

from which we obtain

$$
\bar{S}_{\text{nint}}(\tau) - S_{\text{int}}(\tau) = \mathcal{I}(\tau) \approx (\tau+1)\frac{\overline{\Delta Z^{(\ell)}}(\tau)}{Z_{\text{int}}(\tau)}.
$$

It is worth remarking the dependence of the intertwining on the average variance of single-layer eigenvalues, stressing the fact that it can be used as a measure of topological diversity of layers. In practice, it is desirable to work with the relative intertwining, defined by

$$
\mathcal{I}^{\star}(\tau) = 1 - \frac{S_{\text{int}}(\tau)}{\bar{S}_{\text{nint}}(\tau)}.
$$

Since $S_{\text{int}}(\tau) \leqslant \bar{S}_{\text{nint}}(\tau)$ then $0 \leqslant \bar{S}_{\text{nint}}(\tau) - S_{\text{int}}(\tau) \leqslant \bar{S}_{\text{nint}}(\tau)$, therefore $0 \leqslant \mathcal{I}^{\star}(\tau) \leqslant 1$. For small perturbations, the relative intertwining reduces to

$$
\mathcal{I}^{\star}(\tau) = \frac{(\tau+1)\overline{\frac{\Delta Z^{(\ell)}(\tau)}{Z_{\text{int}}(\tau)}}}{(\tau+1)\frac{1}{L}\sum_{\ell=1}^{L}\frac{Z^{(\ell)}(\tau)-1}{Z^{(\ell)}(\tau)}} = \frac{\overline{\frac{\Delta Z^{(\ell)}(\tau)}{Z_{\text{int}}(\tau)}}}{\frac{1}{L}\sum_{\ell=1}^{L}\frac{Z_{\text{int}}(\tau)+\Delta Z^{(\ell)}(\tau)-1}{Z_{\text{int}}(\tau)+\Delta Z^{(\ell)}(\tau)}}
$$

$$
\approx \frac{\overline{\frac{\Delta Z^{(\ell)}(\tau)}{Z_{\text{int}}(\tau)}}}{\frac{1}{L}\sum_{\ell=1}^{L}\frac{Z_{\text{int}}(\tau)+\Delta Z^{(\ell)}(\tau)-1}{Z_{\text{int}}(\tau)}\left(1 - \frac{\Delta Z^{(\ell)}(\tau)}{Z_{\text{int}}}\right)}
$$

$$
\approx \frac{\overline{\frac{\Delta Z^{(\ell)}(\tau)}{Z_{\text{int}}(\tau)}}}{\frac{1}{L}\sum_{\ell=1}^{L}\frac{Z_{\text{int}}(\tau)+\Delta Z^{(\ell)}(\tau)-1}{Z_{\text{int}}(\tau)}} = \frac{\overline{\Delta Z^{(\ell)}}(\tau)}{Z_{\text{int}}(\tau) + \overline{\Delta Z^{(\ell)}}(\tau) - 1}
$$

$$
= \frac{(N-1)\frac{\tau^2 e^{-\tau}}{2}\overline{\langle(\Delta\lambda^{(\ell)})^2\rangle}}{(N-1)e^{-\tau}\left[1 + \frac{\tau^2}{2}(\langle\lambda^2\rangle - 1) + \frac{\tau^2}{2}\overline{\langle(\Delta\lambda^{(\ell)})^2\rangle}\right]}
$$

$$
= \frac{\frac{\tau^2}{2}\overline{\langle(\Delta\lambda^{(\ell)})^2\rangle}}{1 + \frac{\tau^2}{2}(\langle\lambda^2\rangle - 1) + \frac{\tau^2}{2}\overline{\langle(\Delta\lambda^{(\ell)})^2\rangle}}. \tag{I12}
$$

It is worth noting that the second-order approximation used in this section requires that $\frac{\tau^2}{2}\overline{\langle(\Delta\lambda^{(\ell)})^2\rangle} \ll 1$ and $\frac{\tau^2}{2}(\langle\lambda^2\rangle - 1) \ll 1$. It follows that the relative intertwining can be well approximated by

$$
\mathcal{I}^{\star}(\tau) \approx \frac{\tau^2}{2}\overline{\langle(\Delta\lambda^{(\ell)})^2\rangle}. \tag{I13}
$$

The last relation highlights how the intertwining is expected to increase for increasing deviation of single-layer eigenvalue spectra from the overall average. Additionally, the relative intertwining can be approximated as $\mathcal{I}^*(\tau) \approx \frac{\overline{\Delta Z^{(\ell)}(\tau)}}{Z_{\text{int}}(\tau)+\Delta Z^{(\ell)}(\tau)-1}$. If the perturbation is sufficiently small, i.e., $\overline{\Delta Z^{(\ell)}}(\tau) \ll Z_{\text{int}}(\tau)$, the approximation reduces to $\mathcal{I}^*(\tau) \approx \frac{\overline{\Delta Z^{(\ell)}(\tau)}}{Z_{\text{int}}(\tau)-1}$, which directly relates the intertwining to the partition function of multiplex network.

The integral over time of the relative intertwining is used as a quality function in the main text in order to have a parameter-independent quantification of diversity. If $m = 0, 1, \dots, L-1$ indicates the aggregation step of the functional reduction procedure, we define the quality function by

$$
\mathcal{I}^{\star}(m) = \frac{1}{\tau_{\text{max}} - \tau_c}\int_{\tau_c}^{\tau_{\text{max}}} d\tau \, \mathcal{I}^{\star}(m;\tau), \tag{I14}
$$

where $\tau_c > \log(N-1)$ and τ_{max} can be arbitrarily large. To allow for a meaningful comparison with some transport properties, as shown in the next section, we can set $\tau_{\text{max}} = N$ without loss of generality: this choice corresponds to let random walkers explore the network with a number of steps that is of the order of system's size. Since $\mathcal{I}^{\star}(m, \tau) = f(\tau)\phi(m)$,

i.e., its dependence on variables can be separated, the resulting quality function will not be affected by the integral of $f(\tau)$ at different aggregation steps, therefore we can focus only on the influence of $\phi(m) = \overline{\langle (\Delta \lambda^{(\ell)})^2 \rangle}$.

To better understand the behavior of $\phi(m)$, let us consider the case of a multiplex network consisting of L layers, among which $m_0 < L$ are identical. For sake of simplicity, but without loss of generality, we can assume that the identical layers are the ones labeled by $1, 2, \ldots, m_0$, where nonidentical layers are labeled by $m_0 + 1, \ldots, L$. It follows that

$$\phi(m) = \frac{m_0 - m}{L - m} \langle (\Delta \lambda^{(1)})^2 \rangle$$
$$+ \frac{1}{L - m} \sum_{\ell = m_0 + 1}^{L} \langle (\Delta \lambda^{(\ell)})^2 \rangle$$

and

$$\phi(m + 1) - \phi(m) = \frac{m_0 - L}{(L - m)(L - m - 1)} \langle (\Delta \lambda^{(1)})^2 \rangle$$
$$+ \frac{1}{(L - m)(L - m - 1)} \sum_{\ell = m_0 + 1}^{L} \langle (\Delta \lambda^{(\ell)})^2 \rangle,$$

suggesting that the difference in the quality function between two successive aggregation steps depends on the trade-off between a negative term (the first one, since $m_0 < L$) and a positive term (the second one). This difference is positive if and only if

$$\frac{1}{L - m_0} \sum_{\ell = m_0 + 1}^{L} \langle (\Delta \lambda^{(\ell)})^2 \rangle \geqslant \langle (\Delta \lambda^{(1)})^2 \rangle. \qquad \text{(I15)}$$

The last result provides us with a deep insight on the behavior of the quality function: it increases after aggregating two layers that are identical—or, equivalently, with an eigenvalue spectrum whose variance is identically displaced from the overall average—if and only if the average variance of the eigenvalue spectra of nonidentical layers is larger than the variance of the eigenvalue spectra of identical ones. Roughly speaking, the quality function increases when superimposing two layers which provide redundant information in the multiplex system whereas it decreases when it is not the case.

Since at the latest possible aggregation step, i.e., $m = L - 1$, the intertwining is zero by construction, when redundant layers are present we expect an optimal step m_{opt} to exist in correspondence of the step where topological diversity is maximum.

APPENDIX J: COLLECTIVE COSINE DISTANCE BETWEEN LAYERS

A standard way to quantify dissimilarity between layers is to use Jensen-Shannon entropy divergence. Indicating by $\boldsymbol{\rho}^{(\ell)}(\tau)$ and $\boldsymbol{\rho}^{(\ell')}(\tau)$ be the density matrices corresponding to two layers ℓ and ℓ' of a multiplex network, their Jensen-Shannon divergence is defined by

$$\mathcal{J}[\boldsymbol{\rho}^{(\ell)}(\tau) || \boldsymbol{\rho}^{(\ell')}(\tau)] = S^{(\text{mix})}(\tau) - \tfrac{1}{2} [S^{(\ell)}(\tau) + S^{(\ell')}(\tau)],$$

formally equivalent to the measure widely used in quantum computing. Here, $S^{(\text{mix})}$ is the entropy of the mixture ma-

trix defined by $\boldsymbol{\mu}^{(\ell,\ell')} = [\boldsymbol{\rho}^{(\ell)}(\tau) + \boldsymbol{\rho}^{(\ell')}(\tau)]/2$. The quantity $\sqrt{\mathcal{J}[\boldsymbol{\rho}^{(\ell)}(\tau) || \boldsymbol{\rho}^{(\ell')}(\tau)]}$ defines a metric distance between layers that is a function of diffusion time τ.

However, Jensen-Shannon distance (JSD) might be not reliable in the case of multiplex systems consisting of multiple connected components and isolated nodes across layers, scenarios typical when analyzing empirical systems. Moreover, it depends on τ; for multi-resolution analysis, this is desirable, while it is less desirable when one is not interested in having results at different timescales.

In this section, we introduce a novel distance measure, that we name collective cosine distance (CCD), which is not affected by the above limitations. At variance with JSD, which is an information-theoretic measure, CCD is a geometric measure based on the calculation of the average angular distance between transition vectors. The transition vector $\vec{T}_i$ associated to node i is a column of the transition matrix $\mathbf{T}$, and encodes the probability of jumps from that specific node to any other node in the network. As the role of a certain node in distribution of flow can vary from layer to layer, $\vec{T}_i^{(\ell)}$ and $\vec{T}_i^{(\ell')}$ ($\ell, \ell' = 1, 2, \ldots, L$) can in general be different. Here, we quantify this difference by their angular distance in the Euclidean space $\mathbb{R}^N$. This distance is calculated as

$$\theta_i^{(\ell,\ell')} = \cos^{-1} \left(\frac{\vec{T}_i^{(\ell)} \cdot \vec{T}_i^{(\ell')}}{|\vec{T}_i^{(\ell)}||\vec{T}_i^{(\ell')}|} \right), \qquad \text{(J1)}$$

where $|\cdot|$ denotes the vector norm. A collective variable of interest is the average angular distance of nodes across two arbitrary layers:

$$\Theta^{(\ell,\ell')} = \frac{1}{N} \sum_{i=1}^{N} \theta_i^{(\ell,\ell')}. \qquad \text{(J2)}$$

Note that in case of isolated state nodes, the corresponding transition vector can be taken as zero. The collective cosine distance (CCD) between two layers, accounting for the collective angle, is defined by

$$d(\ell, \ell') = 1 - \cos \left(\Theta^{(\ell,\ell')} \right) \qquad \text{(J3)}$$

and can be used to quantify the pairwise dissimilarity of layers while accounting for dissimilarities of their diffusion pathways.

It is worth remarking that CCD is more convenient than JSD for the goal of this study. From a computational perspective, it is dramatically faster than JSD and avoid hierarchical clustering: therefore, reduction is not based on a heuristics but it can be the result of an exhaustive search through all possible functional couplings. From a theoretical perspective, CCD does not depend on a temporal scale and naturally deals with the presence of isolated nodes and multiple connected components.

APPENDIX K: LAPLACIAN MATRIX IN PRESENCE OF ISOLATED NODES

Empirical multiplex systems usually consist of several isolated state nodes and, sometimes, of disconnected components. The presence of these isolated nodes leads to singularities when modeling the dynamics on top of the network.

To avoid such singularities, a typical workaround is to add self-loops connecting isolated nodes to themselves. However, this mathematical trick boosts the dynamical trapping and consequently, it dramatically alters the resulting dynamics in an unpredictable way. Another possibility is to add virtual links, with extremely small weights, connecting an isolated state node to all state nodes of the corresponding layer, or to a number of them. This workaround provides a more realistic approach to model a diffusion process on top of multiplex networks. Unfortunately, when the number of isolated state nodes is large, a large portion of the flow goes through the virtual links and, consequently, the resulting dynamics in an unpredictable way.

To overcome those issues, we propose an alternative approach to find the right Laplacian matrix of a multiplex network and its layers when an arbitrary number of isolated state nodes is present. To this aim, we first question the physical nature of isolated state nodes. As they have no contribution in any interaction, they should not have any impact on the overall flow as well. More technically, instead of focusing on the structure, we associate suitable transition vectors to the isolated state nodes in order to remove their influence on the overall dynamics.

Mathematically, let us indicate by $\vec{T}_i^{(\ell)}$ the transition vector for node i in layer ℓ, and, if the node is isolated in that layer, let us indicate the vector by $\vec{T}_i^{(\ell,\mathrm{iso})}$. Our goal is to exclude the effects of isolated state nodes on the transition vector of node i. Let us assume that node i is connected in only m out of L layers of the multiplex network: the corresponding multiplex transition vector is given by

$$\vec{T}_i = \frac{1}{L} \sum_{\ell=1}^{L} T_i^{(\ell)} = \frac{1}{L}\left[\sum_{\ell=1}^{m} T_i^{(\ell)} + \sum_{\ell=m+1}^{L} T_i^{(\ell,\mathrm{iso})}\right]$$
$$= \frac{1}{m} \sum_{\ell=1}^{m} T_i^{(\ell)},$$

from which

$$\frac{1}{L} \sum_{i=m+1}^{L} \vec{T}_i^{(\ell,\mathrm{iso})} = \frac{L-m}{mL} \sum_{\ell=1}^{m} \vec{T}_i^{(\ell)}.$$

Assuming the same contribution for all isolated state nodes in their average, $\vec{T}_i^{(\ell,\mathrm{iso})} = \vec{T}_i^{(\ell',\mathrm{iso})}$, one obtains

$$\frac{L-m}{L} \vec{T}_i^{(\ell,\mathrm{iso})} = \frac{L-m}{mL} \sum_{\ell=1}^{m} \vec{T}_i^{(\ell)}, \tag{K1}$$

that simply reads

$$\vec{T}_i^{(\ell,\mathrm{iso})} = \frac{1}{m} \sum_{\ell=1}^{m} \vec{T}_i^{(\ell)}. \tag{K2}$$

Once transition vectors for isolated state nodes are obtained, one can find the transition matrix of layers with isolated nodes, and consequently the transition matrix of the multiplex, as the weighted average of transition matrices of layers.

APPENDIX L: TRANSPORT METRICS

In this section, we introduce some transport metrics that we used in numerical experiments to validate how functional reducibility enhances the transport properties in complex systems.

1. Diffusion time

The eigendecomposition of the propagator reads $e^{-\tau\mathcal{L}} = \sum_{i=1}^{N} \mathbf{C}(i)e^{-\lambda_i\tau}$. The evolution of the probability vector is then

$$\mathbf{p}(\tau) = \sum_{i=1}^{N} \mathbf{p}(0)\mathbf{C}(i)e^{-\tau\lambda_i}. \tag{L1}$$

Since $\lambda_1 = 0$, the first term of the summation is a constant, whereas the other terms decay exponentially fast. The second smallest eigenvalue λ_2 governs the convergence of the process by setting a typical timescale λ_2^{-1}, which is known as *diffusion time*.

2. Average return probability

The average return probability, within a time τ, encodes the probability to return back to the origin of a random walk. It is is given by

$$\mathcal{R}(\tau) = \int_0^\infty d\lambda\, e^{-\tau\lambda} \rho(\lambda), \tag{L2}$$

where $\rho(\lambda)$ is the spectral density of the normalized Laplacian matrix [50].

In order to allow for comparing different networks regardless of a specific temporal scale, we integrate again within the same interval $\Delta\tau = \tau_f - \tau_i$ used for the overall analysis. Finally, we use

$$\mathcal{R} = \frac{1}{\tau_f - \tau_i} \int_{\tau_i}^{\tau_f} d\tau\, \mathcal{R}(\tau). \tag{L3}$$

Note that we perform the same integration for the partition function, to get rid of its dependence on τ.

3. Navigability

One way to define the *navigability* of a multiplex network is to measure its coverage [6], quantifying the efficiency of the walker in discovering the network in terms of the average fraction of distinct nodes visited at least once within τ steps. For connected networks, this fraction increases over time until the random walker explores all nodes of the network. A good approximation for the coverage is given by

$$\Omega(\tau) \approx 1 - \frac{1}{N^2} \sum_{i,j=1}^{N} (1-\delta_{ij}) e^{-C_{ij}(1)\tau - C_{ij}(2)\lambda_2^{-1}}, \tag{L4}$$

where δ_{ij} is the Kronecker delta, λ_2 is the second smallest eigenvalue and $\mathbf{C}(i)$ depends on the ith eigenvector of the normalized Laplacian matrix. To compare two networks in terms of their coverage, we define their navigability by integrating the corresponding coverage function over time and

then normalizing it by the integral vicinity:

$$\mathcal{N} = \frac{1}{\tau_f - \tau_i} \int_{\tau_i}^{\tau_f} d\tau \, \Omega(\tau). \tag{L5}$$

APPENDIX M: RELATING INTERTWINING TO TRANSPORT PROPERTIES

Let us define $\lambda_2^{(\ell)} = \lambda_2 + \Delta\lambda_2^{(\ell)}$, from which we have in the limit of large τ:

$$Z_{\text{int}} \approx 1 + e^{-\tau\lambda_2},$$

$$Z^{(\ell)} \approx 1 + e^{-\tau\lambda_2^{(\ell)}} = 1 + e^{-\tau(\lambda_2 + \Delta\lambda_2^{(\ell)})} \approx Z_{\text{int}} + \Delta Z^{(\ell)},$$

where

$$\Delta Z^{(\ell)} = -\tau e^{-\tau\lambda_2} \Delta\lambda_2^{(\ell)}. \tag{M1}$$

We have already showed that for large τ the fundamental inequality reduces to

$$Z_{\text{int}} \leqslant Z_{\text{int}} + \frac{1}{L} \sum_{\ell=1}^{L} \Delta Z^{(\ell)}, \tag{M2}$$

leading to

$$Z_{\text{int}} \leqslant Z_{\text{int}} - \tau e^{-\tau\lambda_2} \frac{1}{L} \sum_{\ell=1}^{L} \Delta\lambda_2^{(\ell)}, \tag{M3}$$

or, equivalently,

$$0 \leqslant -\tau e^{-\tau\lambda_2} \overline{\Delta\lambda_2^{(\ell)}}, \tag{M4}$$

from which we deduce that $\overline{\Delta\lambda_2^{(\ell)}} \leqslant 0 \leqslant \lambda_2$. The resulting intertwining becomes

$$\mathcal{I}(\tau) = (\tau+1)\frac{\overline{\Delta Z^{(\ell)}}(\tau)}{Z_{\text{int}}(\tau)} = (\tau+1)\frac{-\tau e^{-\tau\lambda_2}\overline{\Delta\lambda_2^{(\ell)}}}{Z_{\text{int}}},$$

and for the relative intertwining, we have

$$\begin{aligned}
\mathcal{I}^{\star}(\tau) &= \frac{\frac{-\tau e^{-\tau\lambda_2}\overline{\Delta\lambda_2^{(\ell)}}}{Z_{\text{int}}}}{\frac{Z_{\text{int}}-1}{Z_{\text{int}}} + \frac{-\tau e^{-\tau\lambda_2}\overline{\Delta\lambda_2^{(\ell)}}}{Z_{\text{int}}}} \\
&= \frac{-\tau\overline{\Delta\lambda_2^{(\ell)}}}{1 - \tau\overline{\Delta\lambda_2^{(\ell)}}} \\
&\approx -\tau\overline{\Delta\lambda_2^{(\ell)}} = -\tau\left(\overline{\lambda_2^{(\ell)}} - \lambda_2\right) = \tau\left(\lambda_2 - \overline{\lambda_2^{(\ell)}}\right).
\end{aligned}$$

The last step highlights the existence of a direct relationship between our quality function for quantifying functional redundancy and the second smallest eigenvalue of the Laplacian matrix, λ_2. It is easy to show that—regardless of the value of τ—the aggregation step $m = m_{\text{opt}}$ where $\mathcal{I}^{\star}(m; \tau)$ is globally
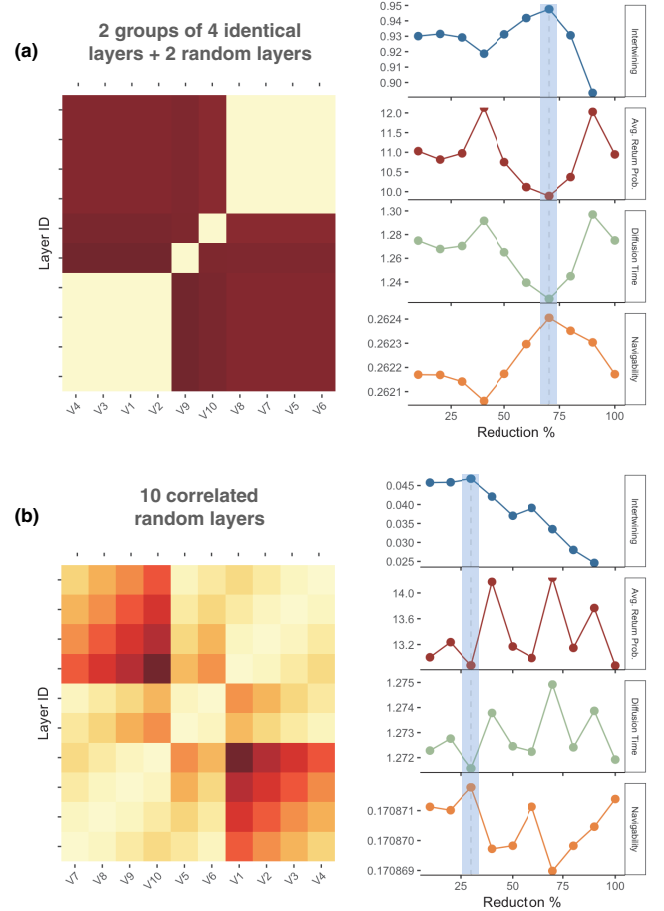

FIG. 7. Functional reduction of additional synthetic systems. See Appendix N for details about the two benchmarks.

maximum corresponds to a minimum in the characteristic diffusion time, quantified by $1/\lambda_2$ and the navigability, which at first order is proportional to $-\lambda_2$.

APPENDIX N: ADDITIONAL SYNTHETIC BENCHMARKS

In Fig. 7(a), we consider a multiplex network with 100 nodes, whose layers are all random networks with wiring probability $p = 0.2$, with the additional constraint that two bunch of layers are exactly the same. Figure 7(b) shows results for a multiplex network of 100 nodes and ten layers, in which the first layer is a random networks with $p = 0.4$, the second layer is a copy of the first but after randomly cutting links with a probability of $4/100$, and so forth so on, to induce correlations among random layers. The functional analysis of these systems is in agreement with our expectation: layers which are exactly the same are coupled together: the resulting groups are noisy and independent networks, which can not be further aggregated.

[1] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, Phys. Rep. **424**, 175 (2006).

[2] X. Gabaix, P. Gopikrishnan, V. Plerou, and H. E. Stanley, Nature (London) **423**, 267 (2003).

[3] M. Bardoscia, S. Battiston, F. Caccioli, and G. Caldarelli, Nat. Commun. **8**, 14416 (2017).

[4] S. Pilosof, M. A. Porter, M. Pascual, and S. Kéfi, Nat. Ecol. Evol. **1**, 0101 (2017).

[5] R. Albert, H. Jeong, and A.-L. Barabási, Nature (London) **406**, 378 (2000).

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

[7] J. Gao, B. Barzel, and A.-L. Barabási, Nature (London) **530**, 307 (2016).

[8] A. Majdandzic, B. Podobnik, S. V. Buldyrev, D. Y. Kenett, S. Havlin, and H. Eugene Stanley, Nat. Phys. **10**, 34 (2013).

[9] A. Majdandzic, L. A. Braunstein, C. Curme, I. Vodenska, S. Levy-Carciente, H. E. Stanley, and S. Havlin, Nat. Commun. **7**, 10850 (2016).

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

[11] J. D. Noh and H. Rieger, Phys. Rev. Lett. **92**, 118701 (2004).

[12] A. Avena-Koenigsberger, X. Yan, A. Kolchinsky, M. van den Heuvel, P. Hagmann, and O. Sporns, PLoS Comput. Biol. **15**, e1006833 (2019).

[13] R. Braden and J. Postel, Requirements for internet gateways, https://www.ietf.org/rfc/rfc1009.txt (1987), [Online; accessed 19-July-2019].

[14] F. Baker, Requirements for ip version 4 routers, https://www.ietf.org/rfc/rfc1812.txt (1995), [Online; accessed 19-July-2019].

[15] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási, Science **327**, 1018 (2010).

[16] G. Smith, R. Wieser, J. Goulding, and D. Barrack, in *2014 IEEE International Conference on Pervasive Computing and Communications (PerCom)* (IEEE, 2014), pp. 88–94.

[17] R. Gallotti, G. Bertagnolli, and M. De Domenico, arXiv:1908.02538.

[18] H. Barbosa, M. Barthelemy, G. Ghoshal, C. R. James, M. Lenormand, T. Louail, R. Menezes, J. J. Ramasco, F. Simini, and M. Tomasini, Phys. Rep. **734**, 1 (2018).

[19] E. Bullmore and O. Sporns, Nat. Rev. Neurosci. **13**, 336 (2012).

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

[21] Y.-Y. Liu, J.-J. Slotine, and A.-L. Barabási, Nature (London) **473**, 167 (2011).

[22] U. Harush and B. Barzel, Nat. Commun. **8**, 2181 (2017).

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

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

[25] J. Gao, S. V. Buldyrev, H. E. Stanley, and S. Havlin, Nat. Phys. **8**, 40 (2012).

[26] 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).

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

[28] F. Radicchi and A. Arenas, Nat. Phys. **9**, 717 (2013).

[29] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, J. Compl. Netw. **2**, 203 (2014).

[30] S. Boccaletti, G. Bianconi, R. Criado, C. I. Del Genio, J. Gómez-Gardenes, M. Romance, I. Sendina-Nadal, Z. Wang, and M. Zanin, Phys. Rep. **544**, 1 (2014).

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

[32] R. Gallotti and M. Barthelemy, Sci. Rep. **4**, 6911 (2014).

[33] F. Radicchi and G. Bianconi, Phys. Rev. X **7**, 011013 (2017).

[34] T. A. Schieber, L. Carpi, A. Díaz-Guilera, P. M. Pardalos, C. Masoller, and M. G. Ravetti, Nat. Commun. **8**, 13928 (2017).

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

[36] L. C. Carpi, T. A. Schieber, P. M. Pardalos, G. Marfany, C. Masoller, A. Díaz-Guilera, and M. G. Ravetti, Sci. Rep. **9**, 4511 (2019).

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

[38] A. Arenas, A. Díaz-Guilera, and C. J. Pérez-Vicente, Phys. Rev. Lett. **96**, 114102 (2006).

[39] D. Gfeller and P. De Los Rios, Phys. Rev. Lett. **99**, 038701 (2007).

[40] F. Krzakala, C. Moore, E. Mossel, J. Neeman, A. Sly, L. Zdeborová, and P. Zhang, PNAS **110**, 20935 (2013).

[41] C. Castellano and R. Pastor-Satorras, Phys. Rev. X **7**, 041024 (2017).

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

[43] U. Fano, Rev. Mod. Phys. **29**, 74 (1957).

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

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

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

[47] A. Cardillo, J. Gómez-Gardenes, M. Zanin, M. Romance, D. Papo, F. Del Pozo, and S. Boccaletti, Sci. Rep. **3**, 1344 (2013).

[48] C. Stark, B.-J. Breitkreutz, T. Reguly, L. Boucher, A. Breitkreutz, and M. Tyers, Nucl. Acids Res. **34**, D535 (2006).

[49] M. Chupeau, O. Bénichou, and R. Voituriez, Nat. Phys. **11**, 844 (2015).

[50] A. Barrat, M. Barthelemy, and A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge University Press, 2008).