Modelling the Process-driven Geometry of Complex Networks

A Thesis Presented to the

Department of Mathematics of the University of Trento

In Partial Fulfillment

of the Requirements for the Degree: Doctor of Philosophy

Author: Giulia Bertagnolli

Supervisor: Prof. Manlio De Domenico

Co-supervisor: Prof. Claudio Agostinelli

June 2022

Table of Contents

Preliminary Content **1**

Introduction **3**

Chapter 1: The diffusion distance **7**
1.1 Basic definitions 7
1.2 Random walks and diffusion on networks 9
1.3 The diffusion distance(s) 12
1.4 Diffusion geometry 15

Chapter 2: The efficiency of parallel communication in weighted networks **23**
2.1 The topological efficiency of networks 24
2.2 Weighted communication efficiencies 26
2.3 The weighted global communication efficiency 30
2.4 Measuring the diffusion efficiency 42

Chapter 3: The diffusion geometry of multilayer networks **45**
3.1 Multilayer networks 45
3.2 Random walks on multilayer networks 48
3.3 The multilayer diffusion distance 52
3.4 Multilayer Diffusion Manifolds 59

Chapter 4: The functional rich-club **69**
4.1 The structural rich club 70
4.2 The functional rich-club 71
4.3 Discussion and conclusion 81

Chapter 5: Conclusion **83**

Appendix A: Supplementary Informations and Materials **87**
A.1 Continuous-time Markov chains 87
A.2 Basic results and proofs about graphs 89
A.3 Multidimensional scaling 91
A.4 Network communication efficiency 91
A.5 Functional rich-club 100

Appendix B: Network science terminology **107**

References **111**

Abstract

Graphs are a great tool for representing complex physical and social systems, where the interactions among many units—from tens of animal species in a food-web, to millions of users in a social network—give rise to emergent, complex system behaviours. In the field of network science this representation, which is usually called a **complex network**, can be complicated at will to better represent the real system under study. For instance, interactions may be directed or may differ in their strength or cost—leading to directed weighted networks—but they may also depend on time, like in temporal networks, or nodes (i.e. the units of the system) may interact in different ways, in which case edge-coloured multi-graphs and multi-layer networks represent better the system.

Besides this rich repertoire of network structures, we have also to consider network dynamics. Dynamics **of** nodes and edges are studied, for instance, in percolation, where nodes and/or edges are removed according to some strategy. Dynamics **on** networks, on the other hand, take place on the nodes or on the edges of the network as, e.g., in a random walk on the set of vertices with probability transitions depending on the network connectivity. Additionally, when dealing with real complex systems, the pair of structure and dynamics cannot be divided so that we speak about the interplay of structure and dynamics.

Building on the foundations of spectral graph theory, of non-linear dimensionality reduction and diffusion maps, and of the recently introduced diffusion distance [Phys. Rev. Lett. **118**, 168301 (2017)] we use the simple yet powerful tool of continuous-time Markov chains on networks to model their process-driven geometry and characterise their functional shape. The original works produced during my PhD are all published (or under submission) and focus on different aspects of networks structure and function, with the common narrative of diffusion geometry. The main results are: (i) the generalisation of the diffusion geometry framework to different types of interconnected systems (from edge-coloured multigraphs to multi-layer networks) and of random walk dynamics [Phys. Rev. E **103**, 042301 (2021)] and (ii) the introduction of new descriptors based on the diffusion geometry to quantify and describe the micro-

(through the *network depth* [J. Complex Netw. **8**, 4 (2020)]), meso- (see the *functional rich-club*) and macro-scale (using statistics of the pairwise distances between the networks node [Comm. Phys. **4**, 125 (2021)]) of complex networks. Besides these methodological articles, I collaborated in the study of the segregation/integration of urban network flows [EPJ Data Sci. **10**, 1 (2021)], in the construction, and its exploratory analysis, of an integrated network of interactions between Sars-CoV-2, human proteins, drugs, and symptoms and diseases [Network and Systems Medicine 130-141 (2020)], and in the first attempt to link the frameworks of *statistical physics of information flow in complex networks* and of *diffusion geometry* [arXiv:2202.09692 (2022)]. Although a full characterisation of the diffusion geometry has not been achieved during the three years of my PhD, this thesis provides a strong and comprehensive basis for future work in this direction.

Preliminary Content

Acknowledgements

I want to thank my supervisors and all the colleagues of the CoMuNe Lab (FBK) for never letting me fall short of support and inspiration. I particularly thank Manlio for pushing me to find, not only elegant mathematical definitions and statements, but also physically-grounded descriptions and explanations of observed patterns in networks, which always correspond to real physical phenomena. I thank Claudio for our chats on Statistics and for instilling in me the longing for bringing more Mathematical Statistics into every field of Science.

I also thank my family: Luca, my siblings, my parents, all my in-laws (Tiziana, Matteo, Stefano, Alice and also the second and third order neighbourhoods in the family network, and my sister-in-law, Maddalena, whom I shared the efforts and satisfaction of the years of our bachelor in Maths). I could never forget all my friends, who, from time to time, may have listened to my outbursts of frustration, annoyance, or over-the-top delight for a kind review (of course by Reviewer 2).

I would also like to thank the University of Trento and the Fondazione Bruno Kessler, all the people in both secretaries and administration offices, because despite the occasional complaints, I am here at the end of this stretch.

Last, but not least, I thank the referees of this thesis professor M. Ángeles Serrano (Dept. of Condensed Matter Physics of the University of Barcelona) and professor R. Lambiotte (Mathematical Institute, Oxford University).

If I still have forgotten someone, it was done unintentionally. Please, do not get *too* bad on me.

Introduction

Networks are a powerful representation of complex systems consisting of many interacting units. These systems are characterised by the emergence of collective behaviours which cannot be simply deduced or predicted by the characteristics and dynamics of its units—the well-known "the whole is more than the sum of its parts" paradigm—and are ubiquitous in the real world. For instance a flock of birds, an ant colony, a school of fish, the Internet, our brain are all self-organised systems which never stop surprising us for the spontaneous and regular patterns they show and the ability of their components to work together as one. Despite the different functions these complex systems carry on, from information processing to routing of people through the globe to life itself, and despite their different scales, when represented as networks all these systems display some common features. In many networks, for instance, we can find a small but significant number of *hubs*, nodes with a very large number of neighbours: the majority of us usually interacts with a few dozens of people in a certain time period, but there are few people having one or two orders of magnitude more interactions. In the World Wide Web there are plenty of pages which are pointed to by few others, but at the same time there are few pages which are addressed to by a lot of other web pages, e.g. national or worldwide COVID-19 dashboards during the current pandemic. This *heterogeneity* in the nodes degree is observed in real-world networks, while it is not in purely random networks, as happens also for the *transitivity*. In a social network, where nodes are people and links represent friendship relations, a high transitivity means that a friend of my friend is, probably, also my friend. Along with social networks, also the brain and the Internet display a large number of triangles and, despite the size of such systems, also a short average distance (path length) between any pair of nodes, while remaining *sparse* networks. These features together characterise the well-known *small-world* phenomenon, which is found across different fields and scales.

Also, the word *scale* occurs frequently when talking about complex systems and networks, usually in expressions like *scale-invariance* or *scale-free*, to indicate a property

of the degree distribution of many real-world networks. In the last decades, however, several works are building on the concept of scale from a more geometrical perspective to provide a richer description of complex networks and their dynamics. Fractal geometry[1] pioneered the geometrical approach to network science, introducing the self-similarity of complex networks, which was later related to, and explained by, their latent hyperbolic geometry[2, 3]. The idea is that nodes are embedded in a latent hyperbolic space and the nearer two nodes are in this space, the more "similar" these nodes are and the more likely they be connected in the network. Strange as it may seem, the notion of *distance* is the common factor of different network geometry frameworks. Graphs and networks are basically defined by their sets of nodes and links, without any relation to a (geometric) space. A distance, called the shortest-path distance or geodesic distance, can be naturally defined counting the number of links in a shortest path connecting two nodes and with this distance, a network already becomes a metric space[4]. Nevertheless, things are more interesting where other network distances are introduced: in the hyperbolic geometry framework geodesics in the latent hyperbolic space correspond, with high probability, to shortest paths in the network[3] and this hidden hyperbolic geometry provided an understanding of self-similarity at a deeper mathematical and physical level. But also dynamical processes taking place on a network induce metrics and, consequently, effective process-driven geometries[5, 6], which have been proven useful for characterising functional communities both in single[5] and multilayer[7] networks, for classifying networks based on the ability of their nodes to communicate[6], for predicting the evolution of contagions[8], and for many other interdisciplinary applications[4].

Network science is, indeed, a very interdisciplinary field. Researchers in social and life sciences have all benefited from the graph theoretical representation of a system adopted in 1736 by Euler to solve the famous Königsberg problem and graph theory evolved from a branch of mathematics into a new branch of science, *network science*, thanks to the contributions coming from different areas, especially statistical physics, ecology and social sciences. In these years of pandemics, not only mathematical modelling and epidemiology have taken advantage of mobility and social network data and tools, but networks have also been central in the study of information (and fake news) spreading and communication strategies[9]. Besides the statistical description of networks and their structure, there is a growing interest on dynamical processes on networks and on the complementary information on networks they provide[10–12]. How can different patterns of pairwise interactions affect the resilience of an infrastructure network to failures or the diffusion of fake news on a social network?

What are the key pairwise interactions that enable an efficient communication in a networked system? How can unexpected *emergent* behaviours of a system, such as ferromagnetism or birds flocking, be traced back to pairwise interactions[13]? Network dynamical processes allow us to relate microscopic structures to global behaviours, to uncover hierarchies and sub-structures at different scales. Moreover, they induce (dis)similarity measures or even distances and richer metric spaces, through which we can study the geometry of the system's functions[4, 5].

The approach of the geometry of network-driven processes—being able to identify latent spaces due to functions of the system— is, of course, particularly appealing when the focus is on the interplay between structure and dynamics. For example, the *communicability distance* in congested urban street networks identified routes with high traffic and congestion hotspots among street intersections, which passed unseen by the usual shortest-path distance and related centrality measures[6, 14]; the *effective distance*[8] reveals the order in the observed, and apparently chaotic, spreading patterns of infectious diseases through mobility networks. Finally, as we will see in the pages of this thesis, the *diffusion distance* allows us to find homogeneous groups among people interacting to reach a consensus[5], but also the connectors between these functional communities[15], and in general, to add to the rich repertoire of well-studied sub-network structures, their complementary functional part.

The aim of this thesis is to generalise the existing framework of diffusion geometry[5] and formalise its key concepts using a metric geometry approach. Besides process-driven geometric communities, using diffusion geometry we can further characterised the functional organisation of complex networks, highlighting central nodes, geometric pathways of information flow and efficient communication strategies.

Thesis outline

Each chapter of this thesis focuses on a specific work carried on in the three years of my PhD. All the works, except for the one presented in Chapter 4, the *functional rich-club*, which is currently under review, are peer-reviewed and published in Q1 journals.

In Chapter 1, after introducing the notation for networks, random walks and distances used throughout this thesis, we present the mathematical formalism of *diffusion distances* and of the *diffusion space*. We will discuss the relationship between the structural and functional characterisation of networks and how they complement each other, and then conclude with the first paper published by the writer during her PhD, namely the Network depth (2019)[15].

The brief introduction of different metric structures over networks of chapter 1

leads us to Chapter 2, which is not strictly related to the diffusion geometry. This chapter is related to two papers[16] and[17], which introduce (the first) and use (the second) a more rigorous measure of communication efficiency for weighted networks, building on the common assumption of shortest-path and least-resistance routing communication. In Chapter 2 I will discuss how this work on the *communication efficiency of network flows* enters the big picture of my PhD project.

Chapter 3 focuses on the generalisation of the diffusion distance framework to multi-layer networks. If the extension of a metric from single- to multi-layer networks is quite straightforward, one has to carefully take into account the typical pathologies of multi-layer networks, e.g. the existence of a node only in some and not in all layers or the different types of inter-layer correlations encoded into the inter-layer connections. Additionally, there are many possibilities for defining random walks on this type of systems: are the links between replicas of the same physical nodes just normal links, i.e. are the probabilities of transitioning between replicas of the same physical node different in nature from the transition probabilities between different state/physical nodes? What is the meaning of a distance between state nodes? Can this be summarised at the physical node level, and does this differ from the distance computed after aggregating all the layers? These are the questions we dealt with in[7].

Chapter 4 presents the last (major) work of these intense three years of PhD: the definition of a functional rich-club. The (structural) rich-club is a group of nodes rich in connections, which tend to be more densely connected among each other than expected by chance, and it is a frequently observed structure or phenomenon in real-world systems, especially in social and brain networks. In ch. 4 we introduce a definition of functional rich-club based on the diffusion distance and argue that a complementary and process-driven approach is mandatory when relating the presence of a rich-club to a particular function of the network, such as integration, rapid spreading of information, or even cognition in brain networks. Our approach to the functional rich-club is strictly diffusion-based, but this framework could be easily generalised to distances based on other dynamics.

Conclusions, ch. 5, summarise the main results of my PhD activity and highlight possible future research directions.

Chapter 1

The diffusion distance

1.1 Basic definitions

A *graph* is a pair $G = (V, E)$ where V is a finite (or sometimes countable) set whose elements $v \in V$ are called vertices or nodes; $E \subseteq V \times V$ is the set of edges (also called links or connections) between pairs of elements of V and it defines a relation, called the *adjacency* relation, between the nodes of G. In this work we deal with finite graphs (or graphs with finite *order*[18]), i.e. $|V| =: N < \infty$, (unless otherwise stated) and we call them networks or complex networks when we want to highlight that they are representations of real-world complex systems. For this reason, graph-theoretic representations of real systems can be enriched adding details about the directionality, strength, cost, and type of the pairwise interactions between the system units. For instance, G can be *directed*, when the adjacency relation is not symmetric, i.e. $v \sim u$ does not imply $u \sim v$, or equivalently in terms of the elements of E, $(u, v) \in E \implies (v, u) \in E$. If G is *undirected*, then an edge is usually indicated by an unordered pair $\{u, v\}$. G can be a *weighted* network $G = (V, E, w)$ where $w : E \to \mathbb{R}$ is a map attributing a weight to each edge; if w is defined over E it is required that $w(e) = w(u, v) \neq 0$, if the support of w is extended to $V \times V$, we set $w(u, v) = 0$ for all $(u, v) \notin E$. The arrangements of connections in G is commonly referred to as its *connectivity* or *structure* and, when G is a weighted network, we may distinguish between the dyadic or topological structure and the weighted structure of G, where the latter accounts also for the distribution of edge weights. The connectivity of G is completely characterised by its $N \times N$ *adjacency matrix* A, whose elements A_{ij} are zero for all pairs of nodes not in E and are equal to 1 otherwise. The weighted structure of a weighted network is described by the *weighted adjacency matrix* W defined by $W_{ij} = w(i, j) = 0$ if $(i, j) \notin E$ and $W_{ij} \neq 0$ for existing edges. Many

measures and descriptors of the network (weighted) structure are then functions of these matrices. Henceforth, let $G = (V, E, ...)$ a possibly directed, weighted network of order $N = |V|$ and *size*[18] $m = |E|$; in the remaining of this section we briefly recall the definitions of mostly used descriptors for G. For more details see[18, 19].

The *degree* of node $i \in V$ is the number of edges which are incident to i, $k_i = \sum_{j=1}^{N} A_{ij}$, if $k_i = 0$ i is said to be an *isolated* node. Similarly, the *strength* of node i in a weighted network is defined by $s_i = \sum_j W_{ij}$. The *degree (strength) sequence* is the row vector $\mathbf{k} = (\mathbf{A1}^T)^T = \mathbf{1}A^T$ (resp. $\mathbf{k} = (\mathbf{W1}^T)^T$), where $\mathbf{1} = (1, \ldots, 1)$ is the vector with all entries equal to 1. For directed networks one has to distinguish between *in-degree* $\mathbf{k}^{\text{in}} = \mathbf{1A}$ and *out-degree* $\mathbf{k}^{\text{out}} = \mathbf{1A}^T$ and the same applies to the strength. Let $\mathbf{D}$ the $N \times N$ diagonal matrix with the degrees on the diagonal, i.e. $D_{ii} = k_{ii}$, then the *combinatorial Laplacian* is $\mathbf{L} = \mathbf{D} - \mathbf{A}$, the *(random walk) normalised Laplacian*[1] is $\tilde{\mathbf{L}} = \mathbf{D}^{-1}(\mathbf{D} - \mathbf{A}) = \mathbf{I} - \mathbf{D}^{-1}\mathbf{A}$ and the *symmetric normalised Laplacian* is $\hat{\mathbf{L}} = \mathbf{D}^{-\frac{1}{2}}(\mathbf{D} - \mathbf{A})\mathbf{D}^{-\frac{1}{2}}$. Of course, the last two normalised Laplacian can be defined only if G does not have *isolated* nodes or adopting the convention $(\mathbf{D}^{-1})_{ii} = \left(\mathbf{D}^{-\frac{1}{2}}\right)_{ii} = 0$ when $k_i = 0$[20].

The adjacency and the various Laplacians are central matrices in spectral graph theory[20] and in network science in general, for their role in defining random walks and diffusion, and metric structures on networks. For instance, consider $(\mathbf{A}^2)_{ii} = \sum_j A_{ij}A_{ji}$: if G is undirected it counts the number of edges attached to node i, the degree of node i; for a generic G it counts the number of walks of length 2 starting and arriving at i and similarly the $ij-$th entry of the $n-$th power of the adjacency matrix counts the number of walks of length n with endpoints i and j. A *walk* between node i and node j is a sequence of nodes $(i = v_0, \ldots, j = v_n)$ such that each pair of successive nodes is connected by an edge. Edges and nodes in a walk can be traversed more than once, if, instead, they are traversed exactly once, the sequence is called a *path*. Observe that a path is a subgraph of G[18], while a walk is not. n, the number of edges in the walk (path) is the *length* of the walk (path). The length of a shortest path between nodes i, j defines a distance on undirected, unweighted networks[2], the *shortest-path distance* denoted here by $D^{\text{sp}}(i, j)$. $D^{\text{sp}}(i, i) = 0$, while if there is no path connecting the pair i, j their shortest-path distance is set to ∞ (and in this case the network is not *connected* and it has more the one *connected component*). The set of nodes V with the shortest-path distance becomes a (finite) metric space, which we indicate as (G, D^{sp}) instead of (V, D^{sp}) since $D^{\text{sp}} = D^{\text{sp}}(E)$. Although this metric structure

[1] We will see the reason of this name in the next section.

[2] On directed networks in general $D^{\text{sp}}(i, j) \neq D^{\text{sp}}(j, i)$ so that D^{sp} is only a quasi-metric

has been used to study, e.g., self-similarity in real systems[4], in many real complex networks the shortest-path distance spans a small range—the so-called small-world phenomenon[21]—so that we cannot rely on this distance alone to study the geometry of networks. Here, come into play dynamical processes on networks and their induced effective geometries of networks. In the next section we introduced random walks and diffusion dynamics on networks and conclude this chapter with the framework of diffusion geometry introduced in 2017 by De Domenico[5], which is the starting point for my PhD research activity.

1.2 Random walks and diffusion on networks

The relation between random walks (as finite space time-homogeneous Markov chains), graphs and electric networks[22–24] is well-known and random walks have been used to tackle both theoretical—e.g. spectral clustering[25] and non-linear dimensionality reduction[26]—and practical problems in different fields of (network) science. In computer science the most famous algorithm based on random walks is PageRank[27], an algorithm for ranking web pages; in neuroscience random walks have been used to model communication dynamics alternative to routing through shortest paths and it is now believed that research in brain communication should focus on dynamic routing strategies which mix random spreading and targeted transmission of information[11, 28]; on social networks diffusion dynamics can model the spread of rumours or approximate consensus and opinion dynamics, see e.g. the voter model[29, 30], or can be used to assess mixing patterns at different scales[31]; statistics and physics have been using random walks and the like since more than a century for sampling[32–34] and model of diffusion and transport dynamics in disordered media[35–38] (and, conversely, networks are used to study branched transport[39]).

We move now to a mathematical introduction to the classical random walk on networks, both in discrete and continuous time and leave for Sec.1.4 the review of key properties and important operators of random walks and their role in the geometric approach to data and network analysis.

Let G be undirected and unweighted, to start with. The matrix $\mathbf{T} = \mathbf{D}^{-1}\mathbf{A}$ can be seen as the transition matrix of a random walk on the network G, i.e. a discrete-time Markov chain on the set of vertices V, with transition probabilities given by the connectivity of G. $T_{ij} = \frac{A_{ij}}{k_i}$ is the probability of going from node i to node j and it is different from 0 only if there is an edge connecting i and j—with the convention $T_{ij} = 0$ whenever $k_i = 0$. Furthermore, from i the random walker has equal probability

of transitioning to each of its adjacent nodes (neighbours).

From this discrete-time random walk we can also define its *continuised* version—a continuous-time Markov chain with exponential holding times with rate $\lambda = 1$ and jumping probabilities given by $\mathbf{T}$[34, 40, 41], see Appx.A.1 for more details. In particular the continuous-time process is generated by the $Q-$matrix $-(\mathbf{I} - \mathbf{T}) = -\tilde{\mathbf{L}}$ and the dynamics is controlled by the forward equation

$$\frac{d}{dt}\mathbf{p}(t) = -\mathbf{p}(t)\tilde{\mathbf{L}}. \tag{1.1}$$

Recall, that a $Q-$matrix[40] Q on a countable set I (in our case, on V) is a square $|I| \times |I|$ matrix with non-negative off-diagonal entries q_{ij}, non-positive diagonal entries q_{ii}, and zero row sums.

Given the initial condition $\mathbf{p}(t = 0) = \mathbf{e}_i$, the canonical vector with components $e_j = 0$ for all $j \neq i$ and $e_i = 1$, which corresponds to the deterministic initial condition with the random walker starting in node i, the solution of eq. (1.1) at time $t > 0$ is $\mathbf{p}(t) = \mathbf{e}_i e^{-t\tilde{\mathbf{L}}}$. The components of the probability vector

$$p_{ij} = \left(e^{-t\tilde{\mathbf{L}}}\right)_{ij} \tag{1.2}$$

can be equivalently seen as the transition probability from node i to node j in the time interval $[0, t]$ or as the probability of finding the random walker (i.e. the process) in node (state) j at time t given that it started in node i at time 0. Let us call $\mathbf{P}(t) = e^{-t\tilde{\mathbf{L}}}$. $\mathbf{P}(t)$ is a stochastic matrix and is given by

$$e^{-t\tilde{\mathbf{L}}} = \sum_{h=0}^{+\infty} \frac{(-t\tilde{\mathbf{L}})^h}{h!}. \tag{1.3}$$

We can simplify the computation of the exponential of the normalised Laplacian matrix using its spectral decomposition, which we derive from those of the symmetric normalised Laplacian. $\hat{\mathbf{L}}$ has a spectrum of real (because of symmetry) non-negative eigenvalues (see[20] or Appx.A.2), that we indicate by $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_N$ with $\{\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_N\}$ being the corresponding orthonormal eigenvectors, i.e. $\hat{\mathbf{L}}\boldsymbol{\phi}_l = \lambda_l \boldsymbol{\phi}_l$ and $\boldsymbol{\phi}_l^T \boldsymbol{\phi}_k = \delta_{lk}$, with δ being the usual Kronecker delta. In matrix form, we can write $\hat{\mathbf{L}} = \boldsymbol{\Phi}\boldsymbol{\Lambda}\boldsymbol{\Phi}^T$. Using the fact that the random walk normalised Laplacian can be written as $\tilde{\mathbf{L}} = \mathbf{D}^{-\frac{1}{2}}\hat{\mathbf{L}}\mathbf{D}^{\frac{1}{2}} = \mathbf{D}^{-\frac{1}{2}}\boldsymbol{\Phi}\boldsymbol{\Lambda}\boldsymbol{\Phi}^T\mathbf{D}^{\frac{1}{2}}$, the propagator of the dynamics reads $e^{-t\tilde{\mathbf{L}}} = \mathbf{D}^{-\frac{1}{2}}\boldsymbol{\Phi}e^{-t\boldsymbol{\Lambda}}\boldsymbol{\Phi}^T\mathbf{D}^{\frac{1}{2}}$. The rows of the matrix $\boldsymbol{\Phi}^T\mathbf{D}^{\frac{1}{2}}$ are the left eigenvectors, i.e. $\boldsymbol{\varphi}_l := \boldsymbol{\phi}_l^T\mathbf{D}^{\frac{1}{2}} = (\varphi_l(i)\sqrt{k_i})_{i=1}^N$, while the columns of $\mathbf{D}^{-\frac{1}{2}}\boldsymbol{\Phi}$ form the set of right

eigenvectors of $\tilde{\mathbf{L}}$, i.e. $\boldsymbol{\psi}_l := \mathbf{D}^{-\frac{1}{2}}\boldsymbol{\phi}_l = \left(\frac{\phi_l(1)}{\sqrt{k_1}}, \ldots, \frac{\phi_l(N)}{\sqrt{k_N}}\right)^T$. These results can also be obtained from the symmetrised transition matrix $\hat{\mathbf{T}} = \mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}$, as in[34].

The left eigenvector plays an important role (assuming the network G is connected) corresponding to the eigenvalue $\lambda = 1$ of the transition matrix $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{T}$, which is (with the proper normalisation) the invariant distribution of the discrete-time random walk. Not surprisingly, the invariant distribution for the continuised random walk (1.1) is the same and is given by the left eigenvector of the $Q-$matrix $-\tilde{\mathbf{L}}$ corresponding to eigenvalue $\lambda_1 = 0$: $0 = \frac{d}{dt}\mathbf{p}(t) = \boldsymbol{\pi}\tilde{\mathbf{L}}$. The invariant distribution is more commonly called *the stationary distribution*, which exists and is unique if the chain is irreducible, i.e. the network is connected (in the discrete-time case the chain should be also aperiodic, i.e. G should also be non-bipartite, while in continuous-time there is no possibility of periodicity, see thms. 3.6.1-3.6.2[40]).

Observe that $\hat{\mathbf{L}}$ is itself a $Q-$matrix and it generates the quantum walk described by the Schrödinger equation (using the common bra-ket notation and with $\mathrm{i} = \sqrt{-1}$ and $\hbar = 1$)

$$\frac{\mathrm{d}}{\mathrm{d}t}|\psi\rangle = -\mathrm{i}\hat{\mathbf{L}}|\psi\rangle$$

which, unlike the classical random walk, does not have a stationary distribution[42, 43].

Also (minus) the combinatorial Laplacian is a $Q-$matrix, with transition rates q_{ij} given directly by the adjacency matrix and rate of leaving each node i given by the respective degree, $-q_{ii} = k_i$. In this case there is a family of independent Poisson processes, one for each edge, having rates given by $-\mathbf{L}$. A random walker in node i moves according to which Poisson process on its out-going edges jumps first, intuitively, to which edge "activates" first. The continuous-time random walk generated by $-\mathbf{L}$ is the edge-centric continuous-time random walk[34], which is also known as the *fluid model*[24] (as we expect from the name, the stationary distribution of the fluid model is uniform) and is important for transport dynamics[44]. The node-centric and edge-centric continuous-time random walks have the same behaviour on regular graphs, but are otherwise related in a non-trivial manner[24]. All these random walks can be easily generalised to weighted networks[20], where edge weights encode the strength of an interaction (in terms of electrical networks weights are conductances instead of resistances).

1.3 The diffusion distance(s)

Let us denote by $\mathbf{p}(t|i) = \mathbf{e}_i e^{-t\tilde{\mathbf{L}}} = \left(e^{-t\tilde{\mathbf{L}}}\right)_{i\cdot}$ the probability vector corresponding to the $i-$th row of the exponential matrix $e^{-t\tilde{\mathbf{L}}} = \mathbf{P}(t)$, which gives the posterior probability distribution of finding the random walker across the network at time t given that the walk started in node i with probability 1. The Euclidean distance between two rows of matrix $\mathbf{P}(t)$ and hence, the distance between two posterior probability distributions corresponding to different starting nodes, induced a distance between the nodes of the network, called the *diffusion distance*[5]:

Definition 1.1 (Diffusion distance)**.** Let $G = (V, E)$ be a network, let $i, j \in V$ be two nodes and let $\mathbf{p}(t|i)$, $\mathbf{p}(t|i)$ be the probability vectors at time $t > 0$ resulting from two continuous-time random walks with the same generator but different starting nodes, i and j respectively. Their diffusion distance at time $t > 0$ is defined by

$$D_t(i, j) = \|\mathbf{p}(t|i) - \mathbf{p}(t|j)\|_2. \tag{1.4}$$

where $\|\cdot\|_2$ indicates the Euclidean distance in $\mathbb{R}^N$.

To be precise this construction yields a family of diffusion distances $(D_t(i, j))_{t>0}$ over time t and, as shown in[5, 45, 46] t serves as a resolution parameter. Intuitively, for increasing time t the Markov chain (or the random walker) can explore larger parts of the network and this is particularly evident if we look at the discrete-time jump chain $\mathbf{p}(n+1) = \mathbf{p}(n)\mathbf{T} = \mathbf{p}(0)\mathbf{T}^n = \mathbf{p}(0)\left(\mathbf{D}^{-1}\mathbf{A}\right)^n$. A known fact is that the $ij-$th entry of the $n-$th power of the adjacency matrix, $\left(\mathbf{A}^n\right)_{ij}$, counts the walks of length n starting in i and ending at j. Integrating D_t over time, to get rid of the dependence on t, yields the persistent functional patterns in networks. Henceforth, we indicate by D_t the distance function over $V \times V$ taking values in $[0, \sqrt{2}]$—since $\|\mathbf{p}(t|i)\|_2 \leq \|\mathbf{p}(t|i)\|_1 = 1$ for all nodes $i \in V$ we have that $D_t(\cdot, \cdot)$ is bounded for all nodes $i, j \in V$—while $\mathbf{D}_t$ denotes the diffusion distance matrix with elements $D_{t;ij} = D_t(i, j)$. $\bar{D}_T$ indicates the average diffusion distance over a time interval $(0 < t_1 < \cdots < t_n = T)$: $\bar{D}_T(i, j) = \frac{1}{T}\sum_{i=1}^{n} D_{t_i}(i, j)$ and $\bar{\mathbf{D}}_T$ is the corresponding matrix. Observe that if the network is connected then as $t \to \infty$ $\mathbf{p}(t) \to \boldsymbol{\pi}$, the stationary distribution independently on the initial distribution and so diffusion distances tend to zero. However, it has been observed[5] that there is a distinction between fast and slow shrinking distances in networks with a modular structure, so that clustering methods on the diffusion distance matrix reveal communities in networks, similarly to spectral graph clustering methods[47] and other algorithms based on random walks[48].

Before moving on with other useful definitions, let us prove that D_t is an actual distance on V. Assume $G = (V, E)$ is an undirected, weighted, and connected network. As we have seen in the previous section, the random walk normalised Laplacian can written in terms of the spectrum and eigenvectors of the symmetric normalised Laplacian $\tilde{\mathbf{L}} = \mathbf{I} - \mathbf{D}^{-1}\mathbf{W} = \mathbf{D}^{-\frac{1}{2}}\hat{\mathbf{L}}\mathbf{D}^{\frac{1}{2}}$ so that also the propagator of the random walk dynamics can be written as

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

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

$$\begin{aligned} D_t^2(i,j) &= \sum_{k=1}^{N} \left[\left(e^{-t\tilde{\mathbf{L}}}\right)_k^i - \left(e^{-t\tilde{\mathbf{L}}}\right)_k^j \right]^2 \\ &= \sum_{k=1}^{N} \left[\sum_{\ell=1}^{N} \frac{\phi^\ell(i)}{\sqrt{s_i}} \phi_\ell(k) \sqrt{s_k} e^{-t\lambda_\ell} - \sum_{\ell=1}^{N} \frac{\phi^\ell(j)}{\sqrt{s_j}} \phi_\ell(k) \sqrt{s_k} e^{-t\lambda_\ell} \right]^2 \\ &= \sum_{k=1}^{N} s_k \sum_{\ell=1}^{N} e^{-2t\lambda_\ell} \left(\frac{\phi^\ell(i)}{\sqrt{s_i}} - \frac{\phi^\ell(j)}{\sqrt{s_j}} \right)^2 . \end{aligned} \tag{1.5}$$

We now assume that $p_t^i(t) = \left(e^{-t\tilde{\mathbf{L}}}\right)_j^i \neq \pi_j$, that is, if and invariant distribution $\boldsymbol{\pi}$ for the Markov chain exists (and it exists since G is connected), t is small enough not to be in equilibrium. Firstly, for any pair of nodes i, j, $D_t(i,j) = D_t(j,i)$ and $D_t(i,j) \geq 0$. $D_t(i,j) = 0$ if $i = j$. The case $\left(e^{-t\tilde{\mathbf{L}}}\right)_k^i = \left(e^{-t\tilde{\mathbf{L}}}\right)_k^j$ for all k is impossible, since it would imply two rows of $e^{-t\tilde{\mathbf{L}}}$ be identical. This, in turn, would imply $\det\left(e^{-t\tilde{\mathbf{L}}}\right) = 0$, but $\det\left(e^{-t\tilde{\mathbf{L}}}\right) = \det\left(e^{-t\boldsymbol{\Lambda}}\right) = \prod_\ell e^{-t\lambda_\ell} > 0$. Hence, $D_t(i,j) = 0$ if and only if $i = j$. The triangular inequality follows from the triangular inequality of the norm. Finally, D_t is an Euclidean distance [49] and the diffusion distance matrix is an Euclidean distance matrix. As a matter of facts, D_t is the Euclidean distance in $\mathbb{R}^N$ between the row vectors forming $e^{-t\tilde{\mathbf{L}}}$.

D_t can also be normalised in different ways, depending on our goal: for graphical purposes it may be useful to normalise D_t dividing it by the maximum over all nodes, i.e. $0 \leq \tilde{D}_t(i,j) = \frac{D_t(i,j)}{\max\limits_{i,j \in V} D_t(i,j)} \leq 1$, or it can be simply divided by $\sqrt{2}$ so that it is

bounded in $[0, 1]$. Interestingly, the average of the squared diffusion distances (ASDD)

$$\mathcal{D}_t^2 = \frac{1}{2N^2} \sum_{i,j=1}^{N} D_t^2(i,j) \tag{1.6}$$

is equal to the sum of the (biased) sample variances of the random vector $\mathbf{p}(t|i)$, $\sum_{i=1}^{N} \sum_{l=1}^{N} (p_{il} - \bar{p}_i)^2$, and so it is a measure of the dispersion of the network. The proof follows from the known equivalence[50]

$$s_X^2 = \mathrm{cov}(X, X) = \frac{1}{2N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} (x_i - x_j)^2$$

for a sample from a random variable X, see the Appx.A.2.

The extensions of this distance to weighted, directed or even more complicated networks, such as multilayers is easily achieved as soon as one can define the transition rules from node to node. For weighted directed networks we just need to replace the dyadic adjacency matrix with its weighted counterpart $\mathbf{W}$ and a node's degree with its out-going strength. In[7], the diffusion distance has been generalised not only to multiplex and interconnected systems, but also to different types of random walk dynamics—e.g. maximal-entropy or PageRank random walks—taking place on the same networked system, as will be seen in ch. 3.

Graph Laplacians, transition matrices and diffusion dynamics have often been related to topology and geometry. For instance, the Laplacian can be decomposed into $\hat{\mathbf{L}} = \mathcal{S}\mathcal{S}^T$, where $\mathcal{S}$ is the normalised incidence matrix of dimension $N \times m$ having entries $S_{ie} = \frac{1}{\sqrt{k_i}}$ and $S_{je} = -\frac{1}{\sqrt{k_j}}$ for edge $e = \{i, j\}$ and zeros otherwise, that maps edges (1-chains or 1-simplexes) into nodes (0-chains or 0-simplexes) $\mathcal{S} : C_1 \to C_0$ (is a boundary operator) and $\mathcal{S}^T$ is its corresponding co-boundary operator $\mathcal{S}^T : C_0 \to C_1$[20, 51]. The spectral decomposition of the various Laplacian matrices are ubiquitous not only in spectral graph theory but also in spectral methods for community detection[46–48, 52], spectral embedding and non-linear dimensionality reduction methods[26, 45, 53], centrality measures[27, 54], and also other network distances, such as the resistance distance[55], the effective distance[23, 56] and Estrada's communicability[6], although this last one is based on the adjacency matrix. This because being able to link discrete objects like graphs that do not possess a natural geometry to actual metric and geometric spaces makes their analysis and our understanding of their properties, generating principles and, ultimately, of their functioning easier. Geometry is among the most basic ways we perceive the world: where are the objects located in

our surrounding, their distances and shapes. Geometry is built in our selves.

1.4 Diffusion geometry

According to the English definition, "geometry" is *concerned with the shape of individual objects, spatial relationships among various objects, and the properties of surrounding space*[57]. Here, we use the expression *diffusion geometry* of a network to indicate how adjacency relations between pairs of nodes translate, through diffusion dynamics, into their mutual positions and distances in space. Through the diffusion geometry we can study the role of individual nodes (centrality), their belonging to different sub-networks, and how each pairwise interaction shapes the whole network. A nice aspect of dynamics on networks, is that they link properties of the network at the micro, meso, and macro scale. Random walks have an additional advantage: through the iterated step of the random walk, we describe the relation between each single node, its neighbours, and the whole in probabilistic terms. Thus, statistics of random walks quantify different properties of networks at different scales: the *expected return time* (or mean recurrence time) to a node i, $m_{ii} = \frac{1}{\pi_i}$ or the mean first passage time m_{ij} from node i to node j are involved in the definition of a number of local and global network dimensions and scaling[34, 58] characterising properties of the network topology or of the diffusion process[59]; the cover time—the time required for a random walker to visit all nodes—or the *coverage*—the fraction of nodes visited at least once in n steps (or time t)—are used to quantify the functionality or navigability of networks[60].

We will now summarise the main results obtained applying the diffusion geometry framework to study the functional shape of complex networks. Here with the term *functional* we mean the shape of networks, in terms of distances, revealed by *functions* of their connectivity, e.g., the spectrum of the adjacency matrix is a structural feature of the system, while the evolution of a density function over a network under the action of a particular transfer operator will reflect functional properties of the system, because it depends both on the structure, through $\mathbf{A}$ but also on the type of dynamics used. The expressions *functional shape* and functional (or process-driven) geometry of a network then refers to the properties of the cloud of points, corresponding to the network nodes, in the diffusion space.

Definition 1.2 (Diffusion space - 1)**.** The vertex set endowed with the diffusion distance, for a fixed value of t, is a metric space (V, D_t). Furthermore, a random walk dynamics maps V to a set of points in space. Indeed, given $i \in V$, $\left(e^{-t\tilde{\mathbf{L}}}\right)_{i\cdot} = \mathbf{p}(t|i)$ is a vector in R^N and the diffusion distance at time t between a pair of nodes i, j is

the Euclidean distance between the corresponding probability vectors. We call this embedding of the network nodes in $\mathbb{R}^N$ through the diffusion distance, the *diffusion space* or *diffusion manifold*.

Sometimes the term *functional* can have a more general meaning, with the consequence that there is ambiguity when talking about function and networks. Take, for instance, the brain represented through a connectome or brain network. Nodes represent neurons or cortical areas and links represent the synapses or fibre bundles connecting the neurons and enabling their communication. In the brain, communication takes the form of an exchange of neuro-physiological signals and what arises from it is cognition. So communication, which is a function (also but not only) of the network structure produces a "phenotypic function": cognition. In the brain along with the connectome, the structural wiring diagram of the brain, there are functional brain networks where links represent statistical correlations between the activity (evolution of signals) in different brain regions and the brain activity depends on the task the subject is doing. The structure, the functional organisation of a network, the functional network and the function of a network are all related in non-trivial ways. Assuming that *communication* is the most basic factor for the emergence of any function in/of a system, we use the expression functional organisation to indicate how the pairwise communication shapes interactions at higher levels (e.g. between groups) and, ultimately, how these allow the system to perform its actual function(s). We will stress this ambiguity across this dissertation and refer the reader to the Appx. B as our attempt to define, with examples, different terms that suffer from it.

To conclude this chapter, we briefly review the role of the diffusion geometry in the uncovering of functional communities and of functionally central nodes.

Functional communities and functional centrality

Community detection is one of the key topics in network analysis, where one tries to uncover particular sub-structures (or meso-scale structures) of a given network. Communities, like clusters in unstructured data, are groups of nodes that establish denser connections between nodes in the same group than with nodes outside and, similarly to traditional data science, there are different methods[61] for detecting *structural* communities in networks, which are, however, out of the scope of this dissertation and will not be discussed here. The functional approach presented in[5] uses the process-driven geometry, or equivalently the network-driven dynamics, to uncover the meso-scale functional organisation of networks.

Figure 1.1 effectively summarises the key points of this approach: firstly, the diffusion distances allow us to have a graphical representation of networks in space which tells us something about the organisation of its nodes. In a) a network with four communities sampled from the Girvan–Newmann[62] model is embedded into $\mathbb{R}^2$ using multidimensional scaling (see Appx. A.3 for details). As time passes, i.e. for larger values of τ, the diffusion distances naturally tend to zero, but, as highlighted in b) and d), with different speeds: inter-communities distances shrink slower than intra-community distances allowing us to uncover different hierarchies of clusters and a varying number of possible modules. Nevertheless, the organisation into four communities is the most persistent one—see panel e)— and also the one maximising the average inter-cluster distances, panel f).

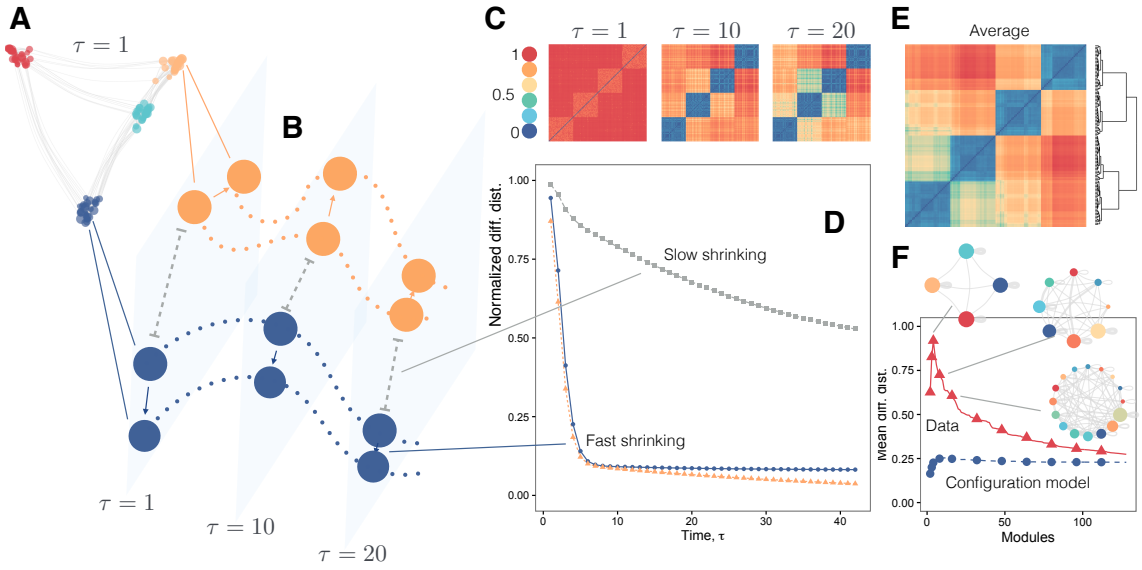

Figure 1.1: Functional communities in the diffusion space. A Girvan–Newmann[62] network with four communities in the diffusion spaces for growing diffusion time τ. In a) its approximated 2D embedding in Euclidean space highlights, together with b) and d) the persistence of the larger inter-communities distances. c) and e) show the diffusion distance matrices at, respectively, different times and the average over time, while in f) all possible hierarchies revealed by the diffusion geometry are shown. The organisation into four communities, i.e. the one reflecting the "ground-truth", maximises the average diffusion distance among communities and can be chosen as the most representative for the given network structure. Figure from[5].

This use of the diffusion distance to characterise the functional cluster organisation of networks is where this metric and the diffusion geometry framework make their first appearance, nevertheless, random walks and other dynamical processes are not strangers to community detection. Other community detection methods, indeed,

are based on dynamical processes[61, 63–65] and random walks in particular[48, 66]. Indirectly, also spectral graph clustering does,s for instance, by taking advantage of the "localisation" of eigenvectors of transfer operators[47, 61, 67] and diffusion propagators[68, 69]. The common idea is that a random walker starting inside a densely connected network will have a small chance to escape the community—i.e. it will be *trapped* in the community[66], which can be rephrased in spectral/matrix terms observing that eigenvectors components with similar values correspond to nodes in the same community (localisation property), both for the adjacency and Laplacian matrices. Recall the classical result of basic graph theory, that we state here for a $k-$regular graph G only for aesthetic reasons: "The multiplicity of the eigenvalue k of the adjacency matrix of G is equal to the number, c, of its connected components". This proposition is proved showing that the c orthogonal eigenvectors of G are the indicator vectors for the vertex sets $V_1, \ldots, V_c$ of the components. Of course this was an extreme case, but it gives the intuition about the underlying mechanism. Also non-linear dimensionality reduction[26], network embedding methods[69], and latent representation learning[70] take advantage of the same property. The main difference and advantage of the approach based on diffusion distances is its geometrical aspect: having an actual metric and an embedding of nodes in a hyperspace of $\mathbb{R}^N$, which have a probabilistic interpretation and also a physical interpretation, since the Laplacian is an averaging operator and can be related to consensus and meta-stable synchronisation dynamics.

At the micro-scale level, i.e. accounting for the importance of individuals units w.r.t. the whole system, we can summarise the information contained in the diffusion distance matrix and get centrality measures. In the spirit of the usual closeness centrality

$$C_i^{\text{clo}} = N \left(\sum_{j=1}^{N} d_{ij} \right)^{-1}$$

we can define the diffusion closeness

$$C^{\text{D-clo}}(i;t) = N \left(\sum_{j=1}^{N} \tilde{D}_t(i,j) \right)^{-1}$$

at a specific time $t > 0$, e.g. the t maximising the inter-community distance, or averaging over time.

A more geometric and statistically robust approach to a diffusion centrality was defined in[15]. The network depth[15] exploits the geometric aspect of the diffusion

distance, and in particular the mapping of the nodes into a cloud of points reflecting the connectivity of the network, to define the network median and, in general, a centre-outward ordering of the nodes, see Fig. 1.2. In non-parametric statistics the generalisation of the median to multivariate distributions is achieved through statistical data depths[71–73], which quantify the depth (centrality) of a point w.r.t. a multivariate probability distribution or, in the sample case, w.r.t. a cloud of points and their empirical distribution.

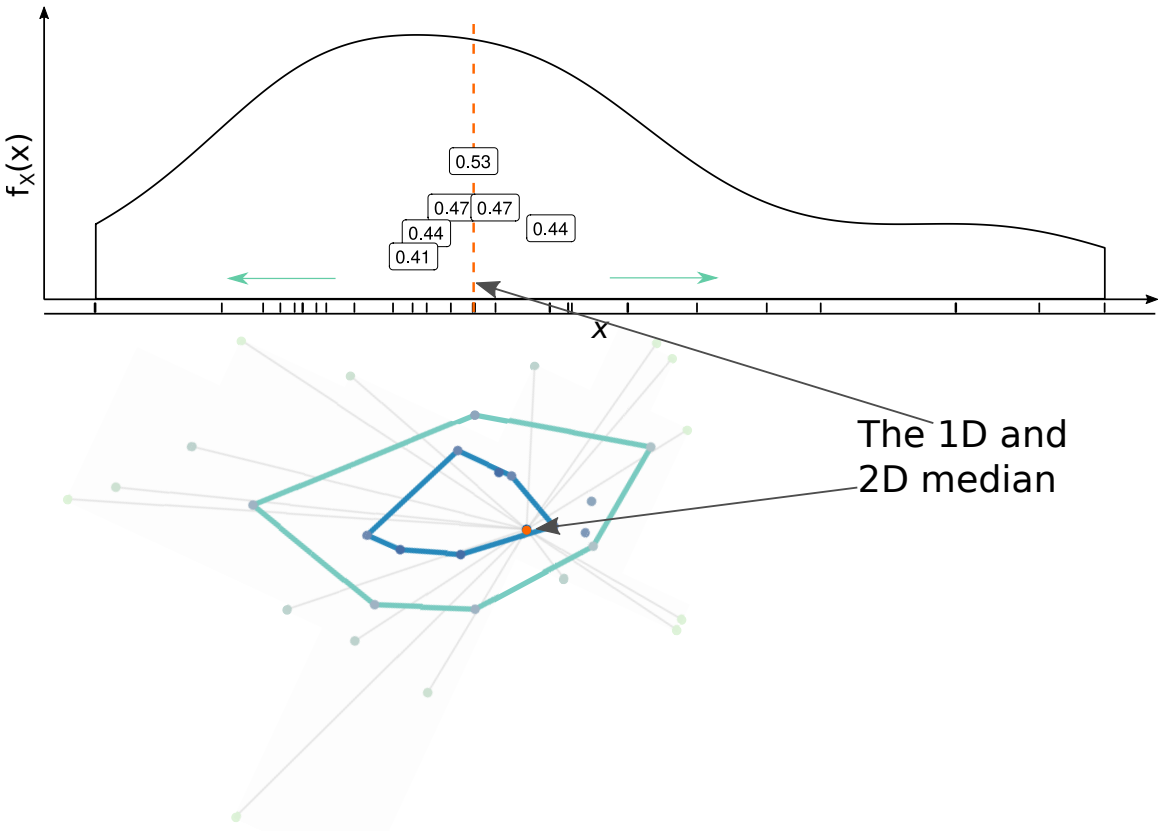

Figure 1.2: Centre-outward ordering of sample points. The usual median and ordering of points from the median to extremal points, in dimension 1 and its extension to a sample of points in $\mathbb{R}^2$ using the half-space Tukey depth[71–73]. Contours and, in higher dimensions, depth-central regions are obtained using the quantiles (order statistics) of the depth.

Applying the same statistical tools to networks embedded into diffusion spaces at different time scales, we found that the deepest nodes are those with a heterogeneous connectivity across the network. In other words, nodes that, instead of having a strong belonging to a community, are well-mixed inside the network structure. For instance, analysing the connectome of the Drosophila[74], Fig. 1.3, we found a high overlap of the network depth, $\mathrm{PTD}(D_t; t, p = 3)$, in the 3D embedding of the diffusion spaces for varying $t > 0$, with Shih's[74] Global Centrality (GC), obtained through a consensus

of a number of different centrality measures, among which the degree, closeness, and betweenness. Such a consensus centrality rewards nodes that are highly central w.r.t. different centralities, that is nodes that have a high number of neighbours and a small shortest-path distance to all other nodes and are in-between a lot of shortest paths connecting pairs of nodes, etc. This means that, although the network depth is a radial measure best suited for finding globally central nodes[15], it does not favour a unique feature of the node, as e.g. its local connectivity alone. Traditionally, formal definitions followed the intuitive idea of the features an important Actor (since many centralities were proposed in the field of social science[75]) should have. Instead, the definition of the network depth is the application of a purely statistical methodology, the generalisation of order relations through depth functions, to networks.

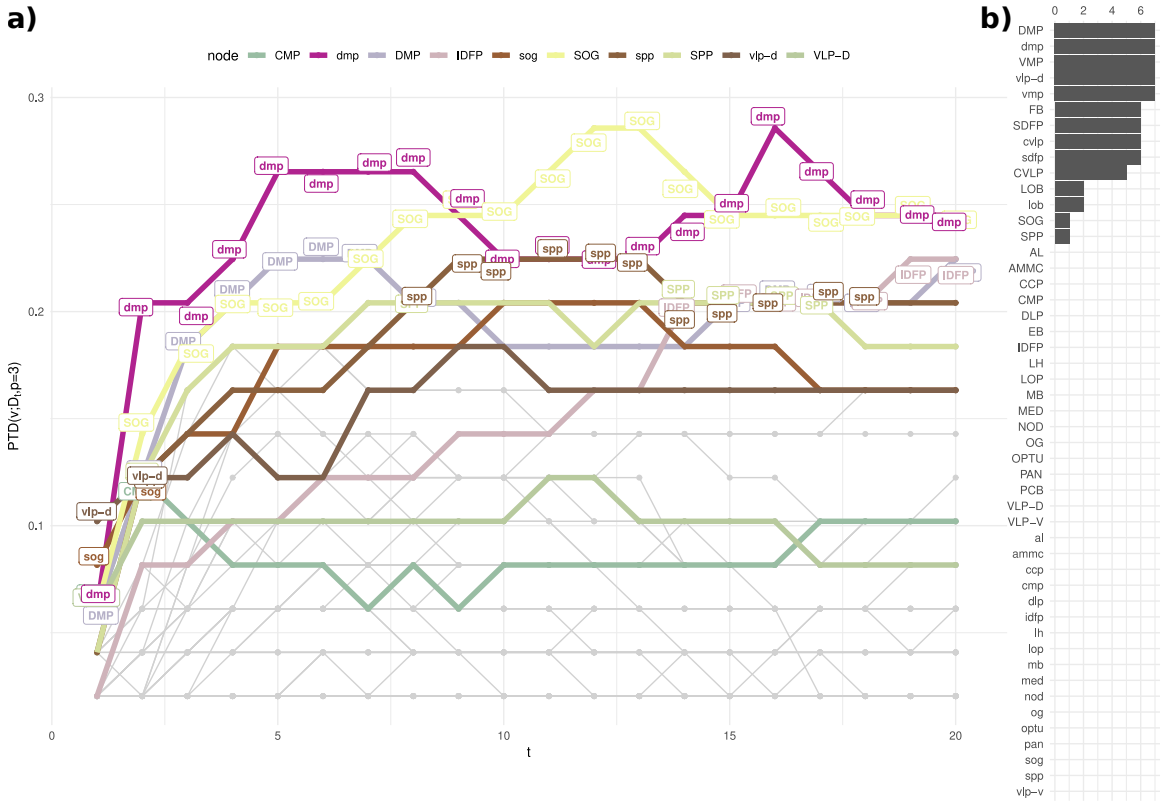

Figure 1.3: Centrality of local processing units (LPUs) in the Drosophila's connectome. *a)* Network depth based on diffusion geometry, top-deepest nodes are selected using the quantiles of the depth. *b)* Shih's[74] Global Centrality (GC) scores for the Drosophila's connectome[74]. The GC is a consensus measure between different centralities. Figure from[15].

The meso- and micro-scale organisation of networks have a key role in shaping other sub-networks or subsets of nodes that have been found to be relevant in real-world networks as it happens, for instance, for the functional rich-club that will be

discussed in detail in ch. 4. Evidently, the functional communities found minimising the inter-group diffusion distances, and the deepest node in the diffusion space depend on the chosen type of random walk and, in general, on the dynamics. However, random walks and diffusion are among the most basic dynamical process that can be defined on networks and can be used as approximations of more complex dynamics as, e.g., meta-stable synchronisation and consensus dynamics[5, 76]. Furthermore, the diffusion distance could be generalised to other dynamics and uncover the different process-driven geometries underlying complex networks, which could then be studied using the same methodologies discussed here.

Networks as metric spaces and kernels on networks

The connectivity among the nodes of a network naturally defines a *similarity* measure over them: nodes that are adjacent have "something" in common, they are related, and are hence nearer in the similarity space induced by the adjacency relation. As we have repeatedly seen in this first chapter, the one-hop connectivity is not the only reservoir for similarity and dis-similarity relations: in an "all-route" communication assumption, the diffusion and communicability distances provide common, recent examples of metric structures on networks, but other examples have been produced also by the natural correspondence between graphs and electric networks. Le us briefly report some results, we address the reader to[41] and references therein for more details.

Consider a connected graph $G = (V, E)$ and let $S \subset V$ and $\phi : V \to \mathbb{R}$ a function such that, for each $i \notin S$ it holds

$$\frac{1}{k_i} \sum_{j \sim i} \phi(j) = \phi(i). \tag{1.7}$$

This is called a harmonic function with set of poles S. Lovasz[41] proposes to consider two (actually three, but we focus on the first two) examples of harmonic functions, which turn out to be strictly related. If $\phi(i)$ denotes the probability that a random walker starting at node i reaches node s before node t, then ϕ is a harmonic function with poles s, t with $\phi(s) = 1$ and $\phi(t) = 0$. Now, from the perspective of electrical networks, each edge in G represents a unit resistance and when an electric current flows through G, entering at s and leaving at t, the same harmonic function with poles s, t can be interpreted as $\phi(i)$ being the voltage of node i. Assume that $S = \{s, t\}$ then the two harmonic functions are the same and, denoting by $R_{s,t}$ the *effective*

resistance between nodes s and t we have that the commute time between nodes s and t is exactly $2|E|R_{st}$ (Thm.4.1[41]), so that commute time can be characterised in terms of electrical network theory:

$$\kappa(s,t) = 2mR_{st}$$

Furthermore, it can be proved that the commute time $\kappa(s,t)$ corresponds to the Euclidean distance between s and t in the Euclidean embedding induced by the group generalized inverse of the Laplace operator $\tilde{\mathbf{L}}$[41, 77]. The same generalized inverse also defines a kernel, i.e. a measure of similarity on the set of nodes.

Another common family of kernels on graphs are the *diffusion kernels*[53, 78], which build on the result that the exponential of any symmetric matrix is symmetric and positive semi-definite. Among this kernels we can find the one defined by the (negative of the) combinatorial or normalised Laplacians, which, on lattices, can be shown to be the finite difference approximation of the continuous Laplacian or Laplace-Beltrami operator on manifolds (involved in the well-known heat equation and *heat kernel*). The heat kernel on graphs has also been used by Chung to define a variant of the PageRank algorithm[79], the heat-kernel PageRank, which is "amenable to various mathematical analyses of the graph"[79] and provides an efficient local partitioning algorithm in terms of Cheeger ratio (i.e., of "bottlenecks" or two "large" subsets of nodes divided by "few" edges)[79].

To conclude, random walks and electrical networks have since long been used to add a metric/geometric structure to graphs and networks, which apparently posses only the metric structure given by the length of shortest paths. On particular topologies, the different distances may coincide, e.g. the effective resistance distance coincides with the shortest-path distance on trees, but they differ in general and the most appropriate metric or embedding to be used, depends on the assumptions of the particular problem, as we will see in the next chapter.

Code release and the `diffudist` R package

Before moving on, a final remark on computational aspects. In these years of *research-apprenticeship*, I learned that it is important to share not only knowledge and results, but also well documented data and code. For this reason the code for evaluating different transition matrices and the induced diffusion distances has been released as an R package, called `diffudist`[80] with a detailed vignette explaining the steps of an example analysis with the package.

Chapter 2

The efficiency of parallel communication in weighted networks

The communication between two individuals, or the exchange of a piece of information between two computers in the Internet, the trade of goods between countries and so on can all be represented as links between nodes in a network. A natural question arising from these different real world systems is whether this trade, exchange, or communication is efficient and to what extent. In network science, where the information available is all and only the one encoded into the sets of nodes and edges with, possibly, their attributes, the most common assumption for quantifying the communication efficiency between two units is the inverse proportionality w.r.t. their *distance* in the system: the further two nodes are in the network, the less efficient is their communication, e.g. due to the higher probability of corruption of a message travelling a long path.

As we have seen in the previous chapters, various metrics can be defined on networks. The natural and most well-known network metric is the length of shortest paths, also called the shortest-path or geodesic distance. As its name says, the shortest-path distance between two nodes is the length, i.e. the number of edges, in a shortest path connecting them. This distance, originally defined for simple unweighted and undirected graphs, can be easily generalised to more complex networks, as long as we are willing to pay the price: for directed networks the shortest-path distance is only a *quasi-metric*, since it is not necessarily symmetric anymore and for weighted networks one first has to figure out what the edge weights encode. As a matter of facts, the weighted shortest-path distance of a pair of nodes is the minimum sum of

lengths (or costs) of edges over all possible paths connecting them. Therefore, if edge weights represent the intensity or the strength of interactions and are hence a measure of proximity and not of distance, we first have to transform edge strengths into edge costs. Furthermore, if the unweighted shortest-path, called *topological shortest-path distance* from now on, has a well-defined scale depending on the system size, the same is not true for its weighted counterpart where weights are, in general, (positive) real numbers. As we will see in this chapter, this complicates things when different systems need to be compared.

The chapter opens with the introduction of the topological, Sec. 2.1, and weighted, Sec. 2.2, descriptors of networks efficiency, discussing both their usefulness and possible issues. In Sec. 2.3 we introduce a new descriptor of weighted efficiency[16], which builds on the topological efficiency measure proposed by Latora and Marchiori[81] and, with a new normalising procedure, overcomes scaling, bias and lacking-robustness issues. In the final section of this chapter we will also outline a possible variation on the communication efficiency, employing diffusion distances, which will remain as a future work.

2.1 The topological efficiency of networks

Let us write in mathematical terms the assumption described above. Let $G = (V, E)$ be an unweighted and possibly directed network with $N = |V|$ nodes, let $i \neq j \in V$ be a pair of nodes and d_{ij} be their shortest-path distance in the network. The pairwise communication efficiency ϵ_{ij} of (i, j) is defined as:

$$\epsilon_{ij} = \frac{1}{d_{ij}}$$

where $\epsilon_{ij} = 0$ if there is no path connecting i and j—remember that in this case $d_{ij} = \infty$ by definition. A global descriptor of the communication efficiency of the network is then defined by averaging over all pairwise efficiencies[81]

Definition 2.1 (Communication efficiency)**.** Let G be a possibly directed network. The global efficiency of parallel network communication is defined as

$$E(G) = \frac{1}{N(N-1)} \sum_{i \neq j \in V} \epsilon_{ij} = \frac{1}{N(N-1)} \sum_{i,j \in V} \frac{1}{d_{ij}} \tag{2.1}$$

Observe that $E(G)$ is bounded in $[0, 1]$, since $\leq d_{ij} \in [1, \infty]$ for all $i \neq j \in V$. Furthermore, $E(G) = 1$ if and only if G is a fully connected network, a clique, with

$E = V \times V$

The concept and descriptor of *efficiency* was introduced in 2001 by Latora and Marchiori[81] to characterise with a unique measure the *small-world effect* observed in many real networks. Traditionally, the small-world effect[21] is defined for unweighted, simple, sparse, and connected networks through two structural indicators: networks with the small-world property have a small average path length, L, and a high clustering coefficient, C. $L = \frac{1}{N(N-1)} \sum_{i \neq j} d_{ij}$ is the average over all shortest-path distances in the network; it is large for $k-$regular lattices ($L^{\text{reg}} \sim \frac{N}{2k}$) and small for random graphs ($L^{\text{rand}} \sim \frac{\ln N}{\ln c}$, where c is the average degree $c = \frac{2|E|}{N}$)[21]. The clustering coefficient C quantifies the transitivity of the adjacency relation in a network, i.e. the probability that two of my neighbours are adjacent to each other[82], and in[21] is defined as the average local clustering coefficient $C = \frac{1}{N} \sum_i C_i$, where the local clustering coefficient of a node i, with k_i neighbours connected by m_i edges, is

$$C_i = \frac{\text{number of triangles connected to vertex } i}{\text{number of triples centered on vertex } i} = \frac{m_i}{k_i \left(\frac{k_i}{2} - 1\right)}.$$

The clustering coefficient, which is bounded in $[0, 1]$, is typically small for random graphs, $C^{\text{rand}} \sim \frac{c}{N}$, while $C^{\text{reg}} \sim \frac{3}{4}$ for regular lattices[21].

$\frac{1}{L}$ is similar to the efficiency as they are both functions of distances, so $E(G)$ can be used as an alternative descriptor for characterising distances/efficiencies in networks, with a major difference: $\frac{1}{L}$ and $E(G)$ involve two distinct means of the shortest-path distance matrix. In the first we find the arithmetic mean, while in the second the harmonic mean, therefore $\frac{1}{L}$ quantifies the efficiency of a sequential communication in the network, while $E(G)$ gives the correct average flow when this flow is exchanged in parallel and concurrently between node pairs[83]. Furthermore, evaluating eq.(2.1) locally over the subgraph G_i, consisting of the first neighbours of node i yields a measure similar to C_i, and by taking their average over all nodes $\frac{1}{N} \sum_i E(G_i)$ provides a global descriptor similar to the clustering coefficient C. To summarise, the efficiency evaluated at different scales can capture both structural properties characterising the small-world phenomenon[81, 83].

Although, by now, the measure of efficiency is purely topological, the parallelism with flow-rates and information exchange is very natural, so it is not surprising that this descriptor was soon generalised to weighted networks[83]. What may seem surprising, is that the weighted descriptor proposed in[83] is rarely adopted and other weighted versions of $E(G)$ exist and are used in applications. In the next section, we

will see why.

2.2 Weighted communication efficiencies

Complex networks are representations of real complex systems, and it is often very important to include edge weights into the network representation[84]. In social networks Granovetter defined the strength of a social tie as "a (probably linear) combination of the amount of time, the emotional intensity, the intimacy (mutual confiding), and the reciprocal services which characterise the tie" and argues that weak ties are of the utmost importance in social integration[85]; in biological networks such as gene regulatory networks and metabolic networks, edge weights encode the intensity of biochemical reactions; in brain networks where edges represent physical connections between neural elements (neurons or cortical areas) edge weights are usually the number of gap junctions and fibre bundles; in infrastructure systems such as transportation and trade networks, the flow of people or goods are important features to consider when assessing the robustness of the system or when modelling the propagation of an epidemic.

As all these examples should have highlighted, edge weights can encode a plethora of different features of real systems so that some assumptions are mandatory when dealing with weighted networks and methods and techniques have to be applied with care to the data at hand. Here, we assume that edge weights are positive real numbers, possibly $w_{ij} \neq w_{ji}$, and that they represent either a flow of a quantity between a pair of nodes—e.g. number of people travelling between two cities, the amount of a good exchanged over a trade network, the number of gap junctions between neuronal elements, the number of emails or face-contacts between two people, the flow of energy or information—or a structural measure of their connection strength, a *conductance*. For instance, here we exclude that in a brain or street network edge weights may encode the physical length of connections. Weighted shortest-paths are then defined as paths with minimum cost or *least resistance* and are found using, e.g., Dijkstra's[86] or Floyd' and Warshall's[87] algorithms. Before these algorithms can be applied to a given a weighted network, edge weights (conductances) w_{ij} have to be first transformed to costs (resistances), usually taking their reciprocal[88–90], $c_{ij} = \frac{1}{w_{ij}}$. Once least-resistance path distances d_{ij}^W are found, a last major difficulty remains, because if we simply plug the weighted distances d_{ij}^W, which vary in $[0, +\infty]$, into eq.(2.1) then $E(G) \in [0, +\infty)$. Therefore, the global efficiency descriptor needs to be rescaled (or normalised) in order to be comparable among different systems.

We henceforth assume that G is a weighted network and we denote by E^{T} its purely topological (global) efficiency[81] computed according to eq.(2.1) discarding edge weights. Observe that in the topological case the term $N(N-1)$ plays the role of a normalising factor, since the parallel sum of shortest-path distances in a clique is exactly equal to $N(N-1)$ and E^{T} naturally lies in $[0,1]$. When shortest-path distances are purely topological, we specify it by d_{ij}^{T}, while $d_{ij} := d_{ij}^{W}$ to simplify the notation. Finally, $E(G)$ is the quantity as in eq.(2.1) with, if not otherwise stated, weighted distances and is, thus, an un-normalised global efficiency descriptor. Each existing re-scaling of $E(G)$ will lead to a normalised efficiency descriptor, denoted with an appropriate superscript.

Let us begin with E^{LM}, the weighted generalization of $E^{\mathrm{T}}(G)$ proposed by Latora and Marchiori in[83]. The idea is to normalise $E(G)$ considering an idealised version of G, G_{ideal}, consisting of a complete network with weighted edges, where the information propagates the most efficiently. Then,

$$\frac{E(G)}{E(G_{\mathrm{ideal}})} = \frac{\frac{1}{N(N-1)} \sum_{i \neq j \in V} d_{ij}^{-1}}{\frac{1}{N(N-1)} \sum_{i \neq j \in V} \ell_{ij}^{-1}} \leq 1. \tag{2.2}$$

Observing that a sufficient condition for (2.2) is

$$0 < \ell_{ij} \leq d_{ij} \quad \forall i \neq j \in V \tag{2.3}$$

defining G_{ideal} reduces to building the matrix $(\ell_{ij})_{i,j}$. They called ℓ_{ij} physical distances, in contrast to shortest-path distances, they assume ℓ_{ij} "to be known even if in the graph there is no edge between i and j", and require that shortest-path distances are computed using "the information contained both in the binary adjacency matrix and in $(\ell_{ij})_{i,j}$"[81]. The matrix $(\ell_{ij})_{i,j}$ is, in every respect, a matrix of ideal connection costs and the condition (2.3) means that each ideal cost should be not larger than the cost of a shortest path.

For spatial networks, such as road, transportation, or brain networks, physical distances may be known—the length of a road, the geographical distance between two locations, the Euclidean distance between two brain regions—, but even in this case we may want to consider other features of the edges which are not easily combined with these physical/geographical distances. For non-spatial systems—such as social and socio-technical systems—the authors propose to define $(\ell_{ij})_{ij}$ as *ad hoc* transformations of edge weights into connection costs. For instance, in a biological network, where w_{ij}

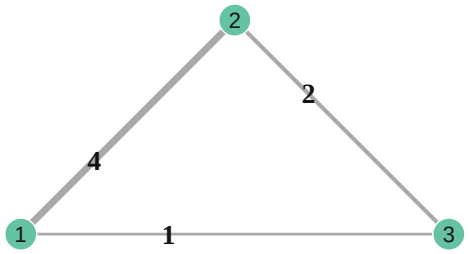

Figure 2.1: Toy network

represents the velocity of chemical reaction along a direct connection between i and j, ℓ_{ij} could be taken as its inverse[81]; if weights $w_{ij} \geq 1$, e.g. when multiple edges are aggregated, one can define $\ell_{ij} = \min\left\{1, \frac{1}{w_{ij}}\right\}$[83], which is the transformation adopted in this work to compute E^{LM}. Unfortunately, also this apparently straightforward procedure hides several issues, e.g. if there is no direct connection between two biochemical units in a connected network, their physical distance is infinite according to the previous definition, while their weighted shortest-path distance will be some positive real number, violating (2.3). Furthermore, in case of real positive weight $w_{ij} \in \mathbb{R}_+$ one cannot take $\ell_{ij} = \min\{1, \frac{1}{w_{ij}}\}$, since this introduces a cut-off on weights smaller than 1. Consider the toy network depicted in Fig. 2.1 consisting of three nodes and three edges, with weighted adjacency matrix:

$$\mathbf{W} = \begin{pmatrix} \cdot & 4 & 1 \\ 4 & \cdot & 2 \\ 1 & 2 & \cdot \end{pmatrix}$$

We can compute physical distances ℓ_{ij} following the suggestions in[83] and shortest-path distances d_{ij} minimizing the sum of costs (i.e. inverse weights)

$$\mathbf{L} = \begin{pmatrix} \cdot & \frac{1}{4} & 1 \\ \frac{1}{4} & \cdot & \frac{1}{2} \\ 1 & \frac{1}{2} & \cdot \end{pmatrix} \qquad \mathbf{D} = \begin{pmatrix} \cdot & \frac{1}{4} & \frac{3}{4} \\ \frac{1}{4} & \cdot & \frac{1}{2} \\ \frac{3}{4} & \frac{1}{2} & \cdot \end{pmatrix}.$$

Observe that the condition (2.3) is not satisfied for $i = 1, j = 3$ and this causes $E^{\mathrm{LM}}(G) = \frac{E(G)}{E(G_{\mathrm{ideal}})} = (22/9)(7/3)^{-1} = 1.047619 > 1$.

This counter-example on the algorithm of[83] and normalising condition (2.3) is not a pathological case: eq.(2.3) is violated whenever the weighted shortest-path between adjacent nodes i, j does not traverse the existing direct link e_{ij}, i.e. $d_{ij} < \frac{1}{w_{ij}}$ and it may often happen in real networks with large heterogeneous weights. See the *Methods* of[16] for additional details.

This difficulty in finding a proper transformation of edge weights into physical distances satisfying eq.(2.3) led researchers[91–94] to adopt another normalised efficiency descriptor in applications, indicated here as E^{MN}. This is the simplest generalization of E^{T} to the weighted case and it, firstly, involves a re-scaling to $[0,1]$ of edge weights, which implies that edge costs and weighted shortest-path distances $d_{ij} \geq 1$. Then, the normalised efficiency can be computed directly by eq.(2.1), since it is already bounded in $[0,1]$.

There are different possible re-scaling transformations of weights, the most common is the max-normalisation, from which the superscript MN: $\tilde{w}_{ij} = \frac{w_{ij}}{\max\limits_{i,j} w_{ij}}$. We can show that $E^{\mathrm{MN}}(G) = \frac{E(G)}{w_{\max}}$. Let $w_{\max}$ be the maximum weight over all edges of a weighted network $G = (V, E)$ and let $SP(i,j)$ be a weighted shortest path between $i, j \in V$. Observe that the max-normalisation of weights does not affect shortest paths, but only their weighted length, i.e. the shortest-path distances

$$\tilde{d}_{ij} = \sum_{n,m \in SP(i,j)} \frac{1}{\tilde{w}_{ij}} = \sum_{n,m \in SP(i,j)} \frac{w_{\max}}{w_{ij}} = w_{\max} d_{ij}.$$

Finally,

$$E^{\mathrm{MN}}(G) = \frac{1}{N(N-1)} \sum_{i \neq j \in V} \tilde{d}_{ij}^{-1} = \frac{1}{w_{\max}} E(G).$$

This fact, may be computationally appealing, but it is definitely not from a statistical point of view: the sample maximum and minimum are the least robust statistics, they are maximally sensitive to outliers. For this reason, E^{MN} may have very wild fluctuations over topologically similar networks but with different maximum weights, which makes this indicator not well suited for comparisons between different systems. To make it clearer, using this descriptor you might not be able to tell if your systems show different global efficiency values because they are characterized by different topologies and interplay between topology and flows, or because their maximum weights are, or are not, outliers to their weights distributions, a very global and extreme feature of the network. Of course, not only the max-normalisation and inverse are available, for instance, in[95] weights are wavelet correlation coefficient between regions in the brain and the cost of the connection between regions i and j is defined as $c_{ij} = 1 - w_{ij} \in [0,2]$. Here the maximum cost, reached for perfect negative correlation, is theoretical and dividing by 2 should not lead to the issues we just outlined. However, this is a particular case and we would like to have a more

general, statistically robust, and physically motivated normalising procedure for the weighted communication efficiency. In[16] we proposed such a normalising procedure, which has also been used in an applied work[17] on urban flows.

2.3 The weighted global communication efficiency

To overcome the issues outlined in the previous section, in our normalising procedure we build G_{ideal} from the weighted network G so that physical distances (i) are not necessarily calculated from metadata or accessible spatial information, (ii) preserve a local feature and so are more robust to outliers and (iii) always fulfil condition (2.3).

Recall that a path is the sequence of vertices in a non-intersecting walk across the network and its weighted length (total cost) is the sum of weights along that path. Weighted shortest-path distances are then computed minimising the sum of the reciprocals of weights[88, 89], which can be seen as costs, over all paths between node pairs. Let us denote by $SP(i,j)$ a weighted and possibly directed shortest-path from i to j. We consider two quantities related to $SP(i,j)$

$$d_{ij} = \sum_{n,m \in SP(i,j)} w_{nm}^{-1} \quad \text{and} \quad \phi_{ij} = \sum_{n,m \in SP(i,j)} w_{nm}$$

d_{ij} is the *total cost* of the path, i.e. the shortest-path distance between i and j, while ϕ_{ij} is the *total flow* along the same path. The matrix $\mathbf{\Phi} = (\phi_{ij})_{i,j \in V}$ represents an artificial connectivity made of short-cuts, where total flows along shortest paths are delivered in one topological step. Finally, G_{ideal} is obtained averaging between the true structure, given by the adjacency matrix $\mathbf{W}$, and the artificial connectivity given by $\mathbf{\Phi}$, i.e., $\mathbf{W}_{\text{ideal}} = \frac{\mathbf{\Phi}+\mathbf{W}}{2}$. Figure 2.2 illustrates the building process of G_{ideal} for a simple graph, consisting of four nodes, labelled from one to four, and four weighted edges, coloured in grey in Fig. 2.2-a). The corresponding cost of each link is indicated by the label of the orange dotted line. The shortest-path distances computed with these costs are the following:

```
     [,1]  [,2]  [,3]  [,4]
[1,]    0   1/4   3/4   2/3
[2,]  1/4     0   1/2 11/12
[3,]  3/4   1/2     0 17/12
[4,]  2/3 11/12 17/12     0
```

Observe that the least-resistance path between nodes 1 and 3 does not pass through

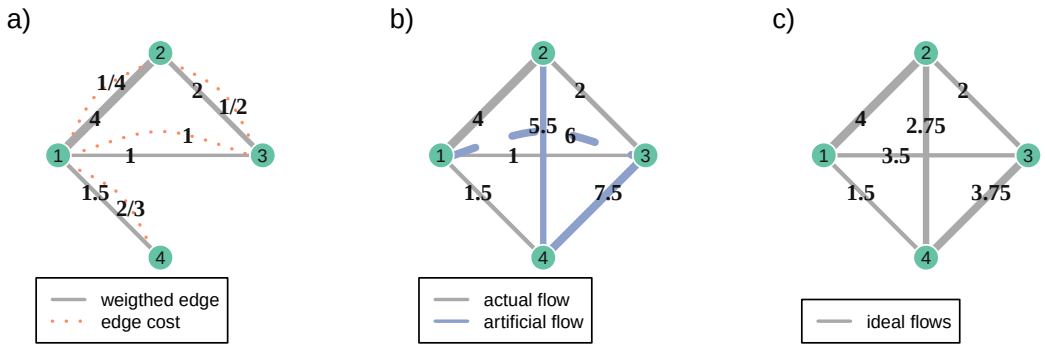

Figure 2.2: Computing ideal flows of a weighted network. **a)** A simple weighted network G with four nodes; links are shown in grey and carry weights; corresponding link costs are shown on dotted orange lines. **b)** Artificial flows are added to G, the dashed edge is the artificial flow through the shortest path, which differs from the direct link $\{1,3\}$, which is topologically short but has a weak flow. **c)** The weighted network G_{ideal}.

the direct link between them, but instead $SP(1,3) = \{1,2,3\}$ and has a total cost of $\frac{3}{4} < 1$. In this case, as shown in panel b) of Fig. 2.2, an artificial link between the pair of nodes is added, although they are already adjacent in G and the artificial flow through this link is $\phi_{13} = 4 + 2 = 6$. Panel c) shows G_{ideal}, where actual and artificial flows are averaged, yielding ideal flows $w_{ij}^{\text{ideal}} = \frac{\phi_{ij}+w_{ij}}{2}$.

Once we have the G_{ideal} corresponding to the network G under study, we can take as physical distances the edge costs in the ideal network, $\ell_{ij} = (w_{ij}^{\text{ideal}})^{-1}$. Firstly, observe that when G is connected, G_{ideal} is completely connected and ℓ_{ij} is finite $\forall i \neq j$. If otherwise G is not connected, G_{ideal} will be disconnected as well. If there is no path between i, j both $\ell_{ij} = d_{ij} = \infty$ and their pairwise efficiency contribute neither to $E(G)$ nor to $E(G_{\text{ideal}})$. We then need to prove that the normalising condition (2.3) is satisfied, and this follows from the Cauchy–Schwarz inequality.

Lemma 2.1. *Let $G = (V, E)$ be a weighted, connected, and possibly directed network with edge weights $w_{ij} \geq 0$ $\forall i, j \in V$ representing interaction intensities or flows, with edge costs defined as $(w_{ij})^{-1}$ and with shortest paths $SP(\cdot,\cdot)$ minimising the sum of costs along any path between node pairs. Let*

$$d_{ij} = \sum_{n,m \in SP(i,j)} w_{nm}^{-1} \quad \text{and} \quad \phi_{ij} = \sum_{n,m \in SP(i,j)} w_{nm}$$

be respectively the total cost of and flow along the shortest path $SP(i,j)$ and let

$$w_{ij}^{ideal} = \frac{\phi_{ij} + w_{ij}}{2}$$

be the idea flow from i to j in the ideal network $G_{idela} = (V, E_{ideal})$ build as described above. Then, for all $i, j \in V$,

$$\ell_{ij} := \frac{1}{w_{ij}^{ideal}} \leq d_{ij}.$$

Proof. For the proof we need the Cauchy–Schwarz (CS) inequality. Let $\mathbf{u}, \mathbf{v}$ be two vectors in an inner product space, the CS inequality reads $|\langle \mathbf{u}, \mathbf{v} \rangle|^2 \leq \langle \mathbf{u}, \mathbf{u} \rangle \cdot \langle \mathbf{v}, \mathbf{v} \rangle$. Taking $\mathbf{u} = \left(\frac{1}{\sqrt{x_1}}, \ldots, \frac{1}{\sqrt{x_n}}\right)$ and $\mathbf{v} = \left(\sqrt{x_1}, \cdots, \sqrt{x_n}\right)$ the inequality becomes

$$\begin{aligned} n^2 = \left(\sum_{i=1}^{n} \frac{\sqrt{x_i}}{\sqrt{x_i}}\right)^2 &\leq \left(\sum_{i=1}^{n} \frac{1}{x_i}\right)\left(\sum_{i=1}^{n} x_i\right) \\ n^2 \left(\sum_{i=1}^{n} x_i\right)^{-1} &\leq \left(\sum_{i=1}^{n} \frac{1}{x_i}\right). \end{aligned} \tag{2.4}$$

Eq.(2.4) states that for non-negative real numbers $x_1, \ldots, x_n$ the inverse of their sum is smaller or equal to the sum of their reciprocals.

Since edge weights are positive, we can apply the inequality (2.4):

$$\left(\sum_{n,m \in SP(i,j)} w_{nm}\right)^{-1} \leq |SP(i,j)|^2 \left(\sum_{n,m \in SP(i,j)} w_{nm}\right)^{-1} \leq \sum_{n,m \in SP(i,j)} w_{nm}^{-1} \tag{2.5}$$

where $|SP(i,j)|$ indicates the number of nodes in the shortest-path. If the shortest-path between i, j coincides with their link (i, j) the number of vertices in the sequence is $|SP(i,j)| = 2$, their shortest-path distance is $d(i,j) = \frac{1}{w_{ij}}$ and $\phi_{ij} = w_{ij}$. Furthermore, $|SP(i,j)| \geq 2$ whenever G is connected, therefore the first inequality is actually strict.

From (2.5) we can derive useful inequalities involving $w_{ij}, \phi_{ij}, d_{ij}$ and ℓ_{ij}:

$$\phi_{ij}^{-1} = \left(\sum_{n,m \in SP(i,j)} w_{nm}\right)^{-1} \leq \sum_{n,m \in SP(i,j)} w_{nm}^{-1} = d_{ij} \tag{2.6}$$

note that if $w_{ij} \neq 0$, it also holds $d_{ij} \leq \frac{1}{w_{ij}}$.

It is also possible to prove that $\phi_{ij} \geq w_{ij}$, $\forall i, j \in V_G$. Indeed, if i, j are not adjacent then $w_{ij} = 0$ but, since G is connected, there is a path between them with $\phi_{ij} > 0$. If instead, they are adjacent, either $\phi_{ij} = w_{ij}$ meaning that the weighted shortest-path coincides with the edge (i, j), or there is a shortest-path going through other vertices, such that $d_{ij} = \sum_{n,m \in SP(i,j)} w_{nm}^{-1} < \frac{1}{w_{ij}}$ and the claim follows from (2.6).

So we have (i) $\phi_{ij}^{-1} \leq d_{ij}$ and (ii) $\phi_{ij} \geq w_{ij}$ for all $i, j \in V$, then (2.5)

$$\ell_{ij} = 2\left(w_{ij} + \phi_{ij}\right)^{-1} \leq 2\left(\sum_{n,m \in SP(i,j)} w_{nm}\right)^{-1} \leq |SP(i,j)|^2 \left(\sum_{n,m \in SP(i,j)} w_{nm}\right)^{-1}$$
$$\leq \sum_{n,m \in SP(i,j)} w_{nm}^{-1} = d_{ij}. \tag{2.7}$$

Again, for a connected network G the strict inequality $\ell_{ij} < d_{ij}$ holds. □

In the lemma 2.1 we assumed G to be connected, but we already observed that if and only if i, j lie in disconnected components also the ideal network will be disconnected as the original one, since $\phi_{ij} = 0$. In this case both $d_{ij} = \frac{1}{\phi_{ij}} = \infty$ and the missing links among disconnected components will not contribute, produce an under-estimation, of the efficiencies of the subgraphs. However, using global indicators for disconnected networks should be avoided.

A final remark, before moving to some application, is on the choice of averaging between the real and the artificial connectivity in G_{ideal}. A stronger option can be to take $\mathbf{W}_{\text{ideal}} = \mathbf{\Phi}$, or to define ϕ_{ij} as the average flow along $SP(i,j)$, instead of the sum. We have the following chain of inequalities:

$$\begin{aligned} d_{ij} &= \sum_{n,m \in SP(i,j)} \frac{1}{w_{nm}} \\ &\geq \frac{1}{\min\{w_{nm} : n, m \in SP(i,j)\}} && (2.8) \\ &\geq \frac{|SP(i,j)|}{\sum_{n,m \in SP(i,j)} w_{nm}} && (2.9) \\ &\geq \frac{1}{\max\{w_{nm} : n, m \in SP(i,j)\}}. && (2.10) \\ &\geq \frac{1}{\phi_{ij}} \end{aligned}$$

So both the minimum, the maximum, and the average weight over the path are valid choices, as well as, the sum (our choice) and the maximum over all edges (the already discussed max-normalisation). Now, when using the sum of flows over paths, we can combine the two sources of information $\mathbf{W}$ and $\mathbf{\Phi}$ through the arithmetic mean, while this strategy is not possible if we define $\phi_{ij}^* = \min\{w_{nm} : n, m \in SP(i,j)\}$ (resp. max) because we cannot prove that $\ell_{ij} = \frac{2}{\phi_{ij}^* + w_{ij}} \leq d_{ij}$, so we should drop $\mathbf{W}$ and simply define $\mathbf{W}_{\text{ideal}} = \mathbf{\Phi}^*$.

In[16] along with the mathematical motivation for proposing our new normalising

procedure, we provided some applications on both synthetic networks generated from models and real networks, comparing our normalised global communication efficiency descriptor, called GCE(G) to the existing descriptors presented in the previous section. The code for evaluating the GCE of complex networks is available as an R package `intsegration`[96] and can be installed from GitHub. The synthetic weighted networks generated in[16], as well as the code for generating them, are also available here[97].

The communication efficiency of synthetic networks

The combination of network topology and flows gives rise to a broad range of systems, so we start with a naive topological structure, a clique of $N = 30$ nodes, and sample the edge weights from two distributions, with varying parameters, to obtain two ensembles with exactly the same topological efficiency $E^{\mathrm{T}}(\cdot) = 1$ but different weighted efficiencies[1], that we measure through $E^{\mathrm{LM}}, E^{\mathrm{MN}}$ and GCE. To simulate homogeneous flows, edge weights are sampled from a Poisson distribution Poisson(λ) with varying λ. Since zero belongs to the support of the distribution, we add one to each sample to keep the complete connectedness of the network. The heterogeneity in the weighted structure is instead modelled with w_{ij} following power-laws(α) with a lower bound $w_{min} = 5$[98]. For each value of the parameters λ and α we take 30 random samples from the respective distribution and generate 30 synthetic weighted networks. Figure 2.3-a) shows their GCEs summarized through boxplots, as a function of λ and α.

All these synthetic networks are topologically equally efficient since they are fully connected, however, accounting for the weights can lead to dramatically different results. The extreme heterogeneity of edge weights, characteristics of power-law distributions with small scaling exponent, strongly reduces the average communication efficiency of the network. Furthermore, as the tails of the weight distributions become lighter, the weighted GCE tends to the topological one. As the parameters λ and α grow, the heterogeneity of weights decreases, since the tailness of the distributions decreases. A measure of the tailness of a distribution is the kurtosis, the standardized central moment of order 4. Usually, one evaluates the kurtosis minus three—the kurtosis of any normal distribution—, which is called the excess kurtosis w.r.t. the normal distribution. For the Poisson distribution, the excess kurtosis is λ^{-1}; for the power-law the excess kurtosis is finite only for $\alpha > 5$ and, for $\alpha > 5$, it decreases as a function of α, tending to 6 as $\alpha \to \infty$ (see the Appx.A.4 for more details). This can

[1]The trivial case, $w_{ij} = w > 0$ constant, leads to $E^{\mathrm{LM}} = E^{\mathrm{MN}} = \mathrm{GCE} = 1$.

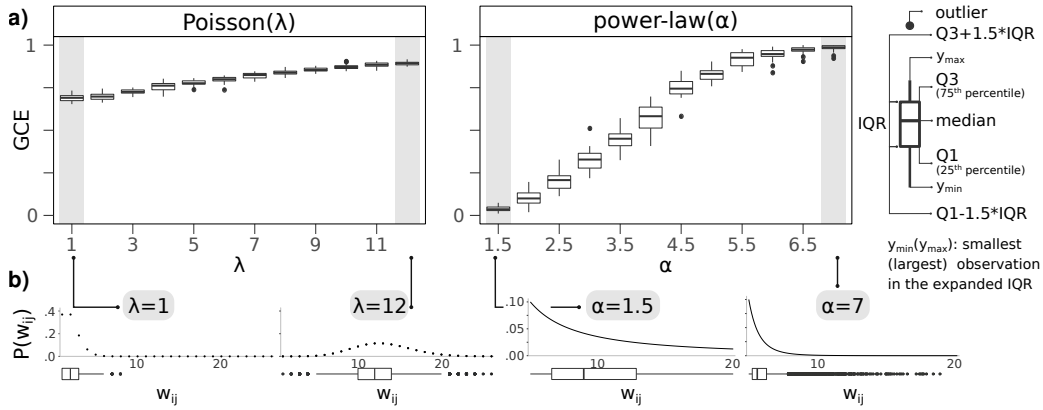

Figure 2.3: Communication efficiency of full networks with homogeneous and heterogeneous flows. **a)** Global communication efficiency (GCE) as a function of the free parameter λ of the Poisson α and power-law distributions from which flows are simulated. The GCE distribution over 30 networks for each parameter value is summarized by the boxplots. In particular the box extends from the first (Q1) to the third (Q3) quartile, the line in the box indicates the median and the whiskers expand the interquartile range (IQR=Q3-Q1) by approx. 1.5 in each direction (see the legend). As α and λ increase the heterogeneity of the weights (kurtosis of the distribution) decreases and the GCEs tend to 1, the efficiency of a fully connected, unweighted network. **b)** The probability mass/density function of edge weights w_{ij} for selected values of λ and α and their boxplots. Figure from[16].

also be seen in the panel b) of Fig. 2.3, where we show the probability mass (resp. density) functions for $\lambda = 1, 12$ and $\alpha = 1.5, 7$.

We next study the interplay between weights heterogeneity and topology through bond percolation, i.e. the targeted attack and removal of the links in our synthetic networks[94, 99]. Edges are removed in decreasing weight order, so that we can trim the tail of the weights distributions, reducing the flow heterogeneity, and move from complete networks to sparser ones.

In Fig. 2.4 we plot the four efficiency quantifiers E^{T}, E^{MN}, E^{LM} and GCE as functions of the fraction f of removed edges and averaged over the 30 random realizations of each model. Shaded areas indicate the standard deviation from the mean. We denote by G_f the damaged network obtained from G removing $f\%$ of its heavier links. G_0 is, topologically, a clique, so $E^{\mathrm{T}}(G_0) = 1$. In $E^{\mathrm{T}}(G_f)$ the denominator is always $N(N-1)$, hence G_f is compared with a clique by definition and $E^{\mathrm{T}}(G_f)$ decreases monotonically. On the other side E^{MN}, E^{LM} (full plots: Figs. A.2-A.3) and GCE use the current flows of G_f to build the corresponding G_f^{ideal} and are, consequently, non-monotone functions of f. This is an advantage for our particular case, for it allows us to compare a series of networks G_f with increasingly sparser and ore different topologies and flows that lose their heterogeneity. For actual

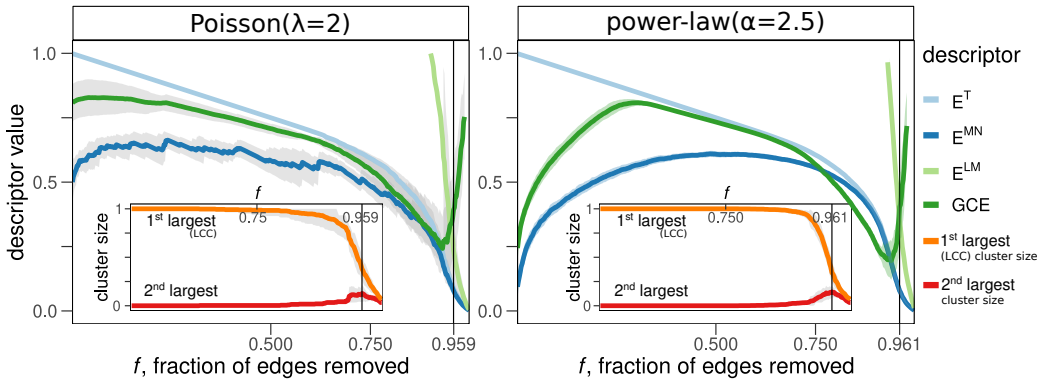

Figure 2.4: Communication efficiency and targeted bond percolation of synthetic networks. From the same network ensembles of Fig. 2.3 we select those corresponding to $\lambda = 2$ and $\alpha = 2.5$ and compare the topological E^{T}, max-normalised E^{MN}, Latora–Marchiori E^{LM} and our GCE efficiency descriptors. Lines represent the average descriptor value over the 30 networks, while shaded areas the standard deviations. The average GCE benefits from the removal of heavy links which forces the network to re-arrange its paths. The insets show the behaviour of the size of the largest and second largest connected component for $f \in [0.5, 1]$. Vertical black lines indicate critical thresholds f_c (corresponding to the maximum of the 2$^{\mathrm{nd}}$ largest component size). Values on the $y-$axis have been cut to the range $[0, 1]$—full plot in the Appx.A.4). Figure from[16].

percolation analysis of networks, however, we propose to modify the GCE, to overcome this possible limitation, as follows:

$$\mathrm{GCE}^*(G_f) = \frac{E(G_f)}{E(G_0^{\mathrm{ideal}})}$$

where G_0^{ideal} is the idealized network corresponding to G_0. Then GCE* is normalized in $[0, 1]$ and it is a monotone decreasing function w.r.t. f, more details in the Appx.A.4.

Back to Fig. 2.4: E^{MN} and GCE behave similarly, although the first has larger fluctuations because at each step the edge with maximum weight is removed and so all flows are rescaled at each step. As expected, there are clear differences in the percolation plots of Poisson and power-law network flows, but in both cases removing the heaviest links produces an increase in the average communication efficiency. In both cases, when the flows become more homogeneous the GCE depends largely on the topology and the two curves GCE and E^{T} are closer. When the network is disrupted—near the critical threshold f_c indicated by the maximum of the second largest component size[100, 101] (see the insets)—the GCE has a break-down point. Here, indeed, we are averaging the efficiencies of many, small, distinct (and maybe individually efficient or "cliquish") disconnected networks, and we already stressed that in this case a global descriptor can fail to give the wished insight. Finally, observe

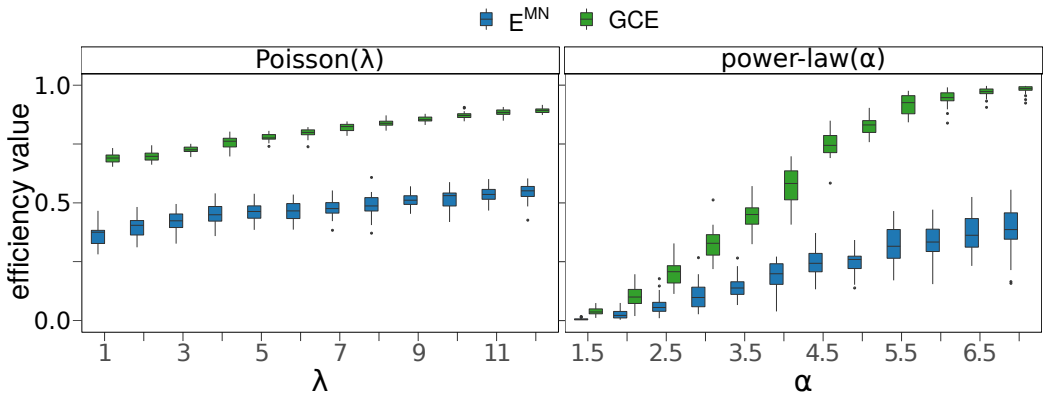

Figure 2.5: Comparison of weighted efficiency descriptors for full networks with homogeneous and heterogeneous flows. As in Fig. 2.3, adding the max-normalized efficiency E^{MN}. For each parameter value the GCE and E^{MN} are evaluated for 30 synthetic networks and their distributions are summarized through boxplots. The GCE converges to 1 as the heterogeneity of the weights distributions decreases (i.e., as λ and α increase), while E^{MN} remains approximately below 0.5. Figure from[16].

that E^{MN} does not increase as much as the GCE when the heaviest 25% of the edges are removed. To better investigate this behaviour, let us look at the boxplots of E^{MN} as functions of λ and α, in Fig. 2.5.

The GCE converges to 1 as the weights become more homogeneous, while E^{MN} remains approximately below 0.5; and these are very specific networks, they are fully connected. What is then the meaning of a descriptor normalized into $[0, 1]$, when the maximum value is so difficult to reach?

We conclude the analysis on synthetic networks softening the constraint we put on the topology and compare the GCE and E^{MN} on synthetic networks generated from models of real-systems, in particular small-world (Watts–Strogatz (WS) model[21]) and scale-free (Barabási–Albert (BA) model[102]) networks. Again we consider 30 random realizations of each topology, each having $N = 256$ nodes, average degree $\langle k \rangle \simeq 12$ and around 5% of all possible edges. Upon these, the edge weights are assigned following the two following rules:

$$w_{ij} = k_i^{\beta} \tag{2.11}$$

$$w_{ij} = e_{ij}^{\beta} \tag{2.12}$$

where $k_i = \sum_{j=1}^{N} A_{ij}$ is the degree of node i, e_{ij} is the (topological) edge-betweenness of the link $\{i, j\}$, and β is a free parameter allowing us to tune the flow structure. The betweenness[62] of an edge $\{i, j\}$ is the number of the shortest paths between any pairs of nodes s, t that go through $\{i, j\}$, here indicated by g_{sijt}, over the total

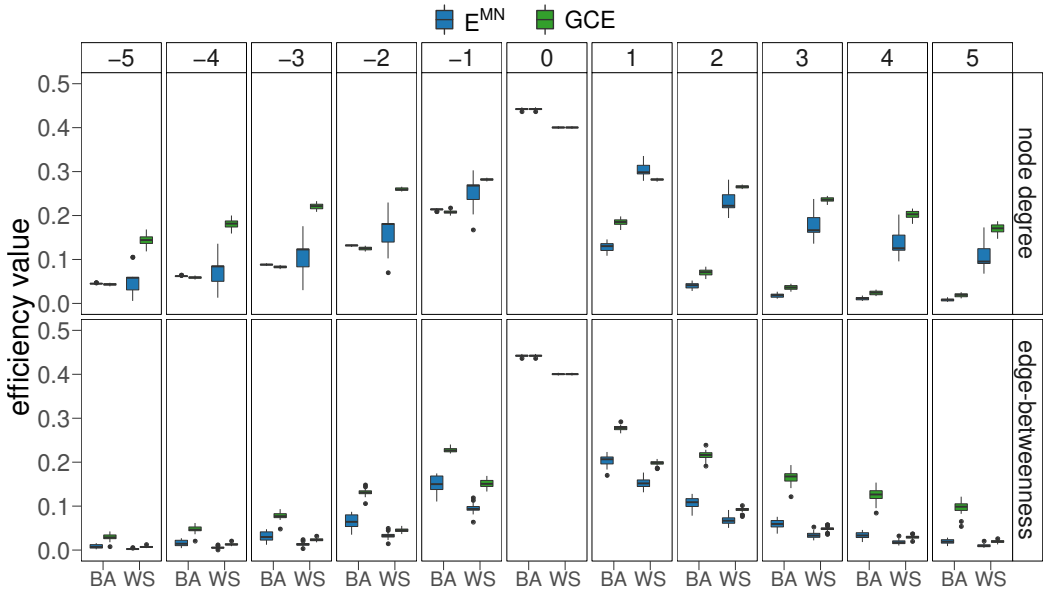

Figure 2.6: Figure from[16].

number of shortest paths between the nodes s and t: $e_{ij} = \sum_{s \neq t} \frac{g_{sijt}}{g_{st}}$.} First, observe that (2.11)[103] generates asymmetric weights; for positive β hubs have strong out-going links, while for negative values of β the intensity of the connections decreases with the degree. The case $\beta = 0$ leads to unweighted systems. Results are summarized in Fig. 2.6. The poor robustness of E^{MN} emerges in the plot: the distributions of E^{MN} over the ensembles are skewed and have greater variance. Therefore, we focus on the GCE for the between models comparison. Topologically, BA networks are slightly more efficient than WS networks, but as soon as weights are introduced the panorama changes: the degree distribution of small-world networks is less heterogeneous w.r.t. the scale-free leading to less heterogeneous weights generated by (2.11) and higher efficiency, independently of the sign of β. On the other side, when links strengths are related to their topological betweenness, BA networks are generally more efficient than WS networks and when $\beta > 0$ the networks are more efficient, because those edges that are very in-between shortest paths are also very efficient. Notice that here, differently from the previous example where all possible links were present, communication paths are unlikely to be able to "re-organize" (i.e. choose a different sequence of edges) in response to weights changes.

The efficiency of real network flows

We use our framework to study the efficiency of four real systems, whose details are summarised in Tab. 2.1. From the FAO worldwide food trade network[104] we selected

the layers of cocoa, coffee, tea, and tobacco. From the migration dataset[105] we selected internal migration flows[106] inside three Asian regions: India, China and Vietnam. From the worldwide air traffic network[8][2] we extracted the traffic in and between Europe and Africa. Finally, we consider the structural connectivity of human brain[107], quantified through diffusion tensor imaging (DTI) and fibre tractography methods.

Table 2.1: Real-world flow networks and corresponding scales. $|V|$ indicates the number of nodes and $|E|$ the number of edges in the networks. Multiple edges have been aggregated and loops removed.

Dataset		$\|V\|$	$\|E\|$	Ref.
FAO	cocoa	159	2,081	[104]
	coffee	184	7,760	
	tea	172	3,297	
	tobacco	183	3,623	
Migrations	China	30	870	[105, 106]
	India	32	992	
	Vietnam	63	3,906	
Transportation	airports	299	12,919	[8]
Biological	human brain	188	10,836	[107]

These real networks have different properties, among which edge density and weight distribution, as shown in Fig. 2.7. The three statistics building up the boxes of the boxplots are the same as before, i.e. first quartile, median (line in the box), and third quartile, while here the whiskers extend from the minimum to the maximum of the distribution to highlight the presence of extreme outliers or very heavy tails. Based on the results of our analysis of synthetic networks and on previous studies[83, 94], we expect the weighted efficiency of these real networks to be smaller than their topological efficiency. Thanks to our normalising procedure, which can be applied unchanged to all networks, it is legitimate to compare the weighted efficiencies of diverse systems, and we expect the trade networks to have the smallest efficiency. As a matter of fact, observing the boxplots in Fig. 2.7 we can see that the distributions of the network flows in the trade networks are highly heterogeneous.

As for the synthetic networks we proceed with a percolation analysis, removing

[2]We thank prof. Dirk Brockamnn for providing us with the worldwide air-transportation flow data.

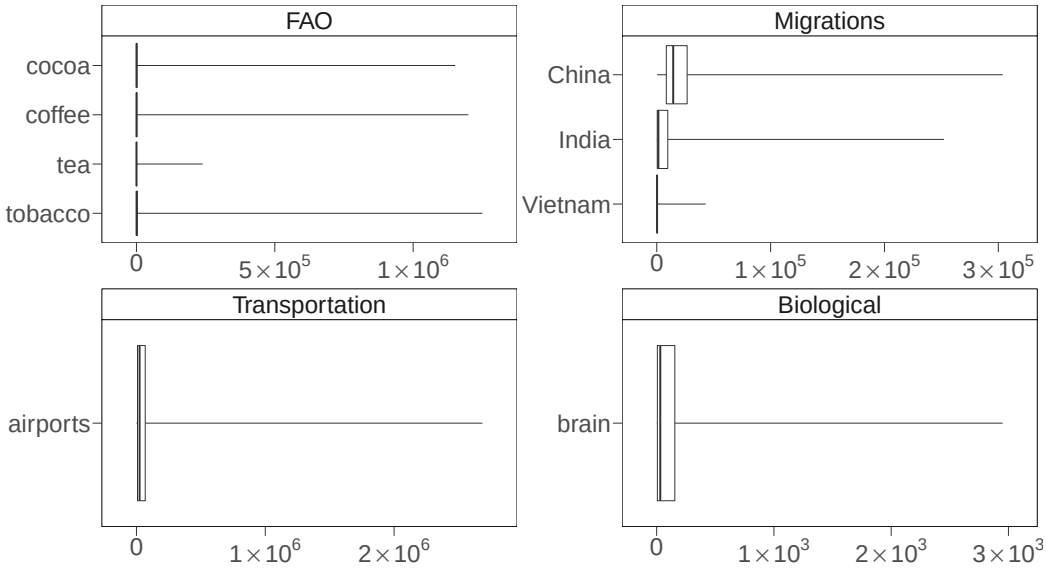

Figure 2.7: Boxplots of edge weights of selected real-world networks. Heterogeneity and scale of flows vary across the datasets.

Table 2.2: Maximum size of the second largest connected component (LCC) of each real network and the corresponding fraction f of removed edges.

f, fr. removed edges	network	nodes in 2-nd LCC
0.978	tea	13
0.434	Vietnam	1
0.989	airports	26
0.995	brain	12

heavy edges from each real network. Figure 2.8 shows the curves corresponding to $E^{\mathrm{T}}(G_f)$ and $\mathrm{GCE}(G_f)$, while Fig. A.7 in the Appx. A.4 shows also $E^{\mathrm{MN}}(G_f)$. Informed by the knowledge gathered from the analysis of synthetic networks, we take a look at the maximum size of the second largest connected component (LCC), which indicates the critical thresholds f_c after which the network is too damaged for the evaluation of a global indicator to make sense. As we can see from Tab. 2.2 the LCC of internal migrations in Vietnam loses one node when 43.3% of its heaviest edges are removed, and also afterwards only isolated nodes detach from the LCC. In Fig. 2.8 we perform the percolation until 99% of the edges in each network are removed, which is near the f_c of almost all four networks. Independently of the system, ignoring the network flows leads to an overestimation of the average efficiency, especially when flows are highly heterogeneous. The network of internal migration, is the most efficient, but it also has the highest cost, being a clique. The tea trade network is the most inefficient. Finally, the brain and the airports network

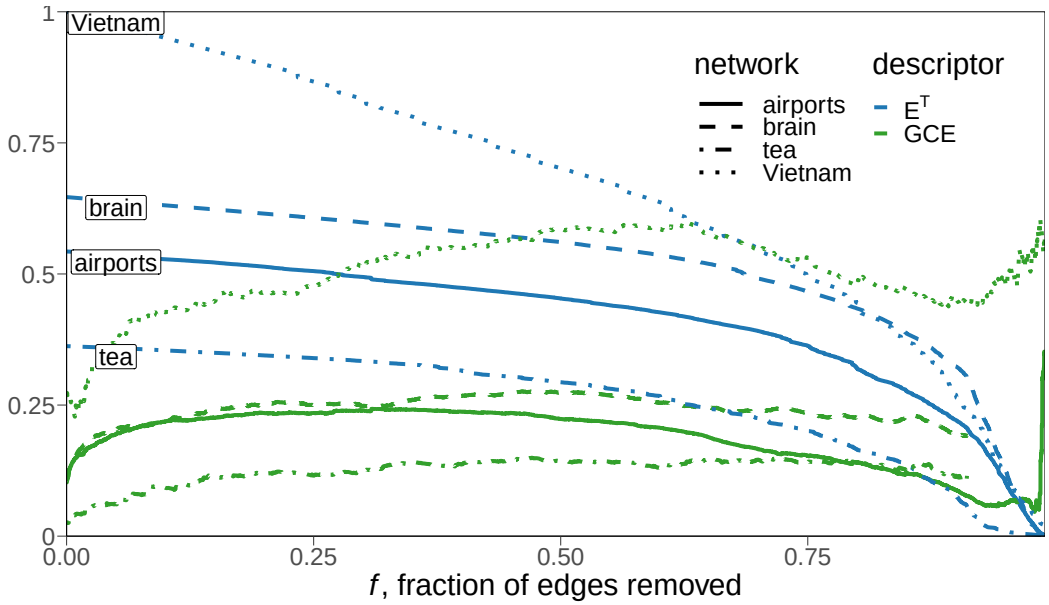

Figure 2.8: Targeted bond percolation of real interconnected systems. The descriptors are the topological efficiency E^{T} and the global communication efficiency GCE of real networks where edges are removed in decreasing order according to their weight. The networks are: The tea trade network, the internal migrations in Vietnam, the air traffic in Europe and Africa and between them, and the human brain network.

have similar GCEs until the first 25% of their edges are removed, with the brain remaining afterwards more efficient w.r.t the reduced flows. Observe that the total flow could be restored, while keeping a specific efficiency value, re-distributing the removed flow on the remaining links. In general, removing those edges monopolizing shortest paths forces their reallocation, inducing an increase of the global weighted efficiency. To summarise, the global communication efficiency, GCE, is based on the definition topological efficiency descriptor, and it shares its assumptions—(i) parallel communication and (ii) routing communication, i.e. through shortest paths. It differs from the topological efficiency since it accounts for edge weights encoding flows and, in general, interaction intensities. The edge weights are accounted for through weighted shortest-path (or least-resistance paths) distances and this introduces the need for a normalisation in the spirit of[81, 83], i.e. through the construction of an idealised version of the network under study. However, the GCE is different from the descriptor in[83] and also from other weighted efficiency descriptors for its normalising procedure, which does not require any metadata, is physically grounded—idea of total flows delivered in one topological step—and results to be also statistically more robust and mathematically correct according to the normalising condition given by Latora and Marchiori in[81, 83]. Using our definition of flow network efficiency, the results

presented in this chapter indicate that one can achieve a desired level of efficiency by wisely rearranging the flows over the network, instead of altering the underlying topology. This result is relevant for practical applications, since it is not always guaranteed that one can rewire or dramatically change with other interventions the network connectivity, since it could be economically or energetically expensive.

We now briefly see how we could weaken the assumptions, in particular (ii), to obtain a measure of network efficiency under other communication paradigms, such as broadcasting or diffusion.

2.4 Measuring the diffusion efficiency

Before turning to the actual concept of diffusion efficiency, we should discuss the motivations behind it, and we start focussing on communication strategies in the brain.

$E(G)$ is a measure of parallel communication efficiency assuming the communication takes place on shortest paths. According to the Merrian-Webster dictionary, communication is *a process by which information is exchanged between individuals through a common system of symbols, signs, or behaviour, also: exchange of information.* So we will use communication and information exchange as synonyms. Now, information can be exchanged in different ways and each information strategy has assumptions and implications. For instance, exchanging information through shortest paths presumes the knowledge of the shortest paths connecting a pair of nodes and hence a complete knowledge of the network structure. If this is the case, this routing communication is very effective: information travels from sender to target and the probability of message corruption is minimised. Instead, if the sender does not have the global knowledge of the network and its shortest paths, an alternative communication strategy is simply passing the message to a random neighbour and wait until the message in delivered. In this case of diffusion the amount of initial information required is limited to the first neighbourhood, but the message will travel longer and possibly redundant (self-intersecting) paths. Routing and diffusion represent the two extremes, but of course, one can define mixed or biased communication strategies and assessing their efficiency with a, possibly common, descriptor is very desirable. Mind that this is not a purely theoretical exercise.

In the brain, communication takes the form of an exchange of neuro-physiological signals and what arises from it is cognition. For this reason, understanding how the pairwise connectivity and communication shape the functionality of the single

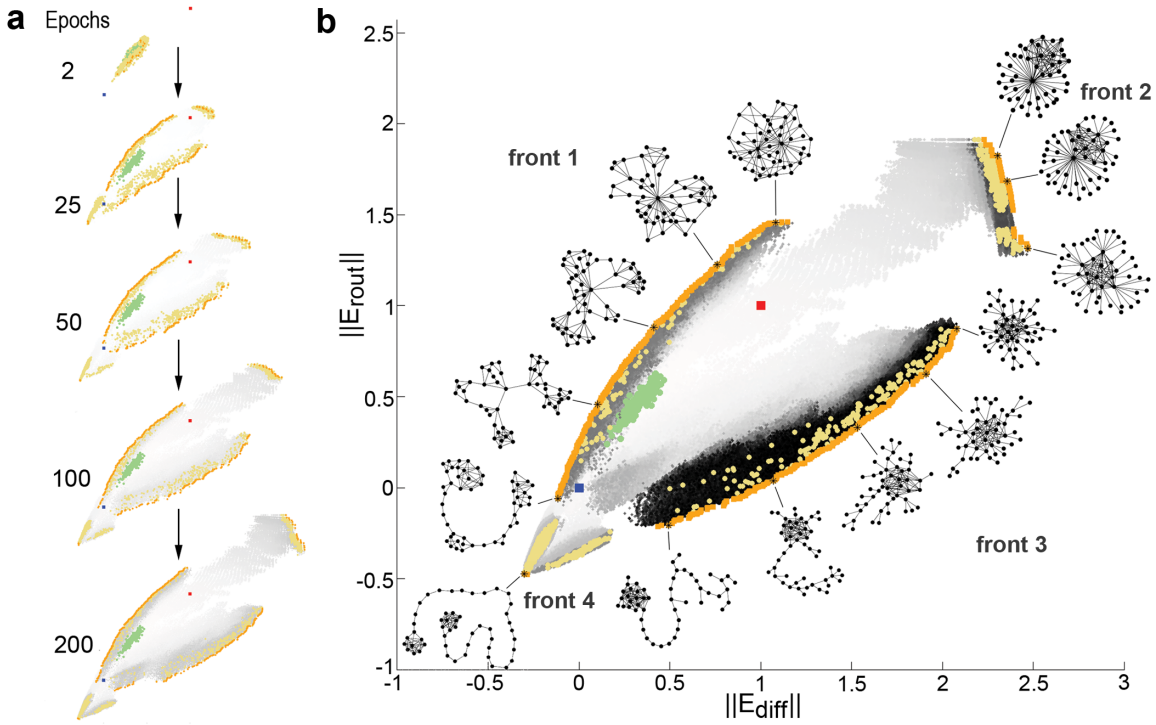

Figure 2.9: Morphospace of communication efficiency of complex networks. Here E_{rout} indicates the original efficiency $E(G)$ of[81] for unweighted networks and both efficiency descriptors are rescaled by $\|E_{\cdot}\| = \frac{E_{\cdot} - \langle E_{\cdot}^{\text{lattice}} \rangle}{\langle E_{\cdot}^{\text{random}} \rangle - \langle E_{\cdot}^{\text{lattice}} \rangle}$. Figure from[109].

neurons and of the whole brain is a major goal in network neuroscience[108]. Without going into much detail, some key-points are now ascertained (see[28] and references therein): (i) the observed near-minimal path length in the brain is achieved at a cost in the brain and not simply as a consequence of the spatial constraints, hence shortest paths have an important role for the brain activity. (ii) Communication in the brain cannot take place exclusively along shortest-paths, because, beyond the already said limitations, this would exclude a large portion of connections to participate in the communication process and also expose the network to vulnerabilities, delays, and congestions. (iii) An entire communication spectrum should be considered and common statistics—like betweenness and closeness centrality and the rich-club—which are used to characterise the brain and are common through multiple species and organisms, should be generalised to encompass other communication strategies.

A first step toward the generalisation of the communication efficiency has been already done[28, 109] and here we are particularly interested in the diffusion efficiency descriptor introduced in[109], which is based on mean first-passage times t_{ij} of a

classical random walk on undirected connected networks:

$$E_{\text{diff}} = \frac{1}{N(N-1)} \sum_{i \neq j} \frac{1}{t_{ij}}.$$

The authors also explored what they called the *morphospace of communication efficiency in complex networks*[109], shown in 2.9, defined by the routing and diffusion efficiencies and explored through a multi-objective optimisation algorithm. Interestingly, the four fronts they found show networks with clear topological features: front 2 consists of chain-like networks and front 4 of star-like networks and represent the regimes where both diffusion and routing efficiency are either low (2) or high (4). The remaining fronts, representing different transitions from 2 to 4 and the regimes where networks are efficient either at diffusing or at routing, are characterised by a modular structure with bridges (front 3) and by a core-periphery structure (front 4) However, this approach has some limitations, in my opinion. Firstly, we can observe that E_{diff} lives on a completely different scale than E_{rout} (see also Fig. 2 of[109]) and a more detailed analysis on the upper-bound of E_{rout} and for which networks it is achieved is mandatory. Secondly, also the scaling of the efficiency descriptors (through comparison with lattices and random networks) should be examined further.

To conclude, there is interest in the network science community, and especially in the neuroscience and social-mobility[17] groups, for measures of efficiency which go beyond the common assumption of communication following shortest-paths. A common approach based on distances, which could include routing, diffusion, other communication strategies lying in between and even other dynamics, would be beneficial not only at the theoretical level but especially for applications.

Chapter 3

The diffusion geometry of multilayer networks

Many real complex systems are characterised by units interacting not only with different intensities, but also in different ways. If the intensity of a pairwise interaction can be encoded into the weight of the corresponding edge, the types of interactions have been traditionally represented through different edge colours, using the edge-coloured graph (or multi-graph) representation of a system[110]. Refining our representations is of the utmost importance when dealing with real data and systems of increasing complexity (a society, the network of connections in the human nervous system or the world-wide food trade network), where, often, the relationships and correlations between the multiple types of interactions matter. The framework of multilayer networks[111] enables the representation of different coupling structures between colours—now layers—in a precise and mathematically grounded[112] way, providing a unified framework for the analysis of multidimensional systems.

3.1 Multilayer networks

A multilayer network may represent a multidimensional system with multiple sub-systems, which may or may not be common to all layers of connectivity. These sub-systems must, however, share some common units, otherwise we could represent the whole system simply as a network of networks. Here we will then assume that multilayer networks are defined on a set V of nodes which are common to all layers. This may seem a strong assumption, especially because in real-world networks it often happens that some nodes are present only in some layers, but in this case we will represent them as isolated vertices in the layers where they are absent, i.e. where they

do not have any interaction. Observe that the formulation of multilayer networks[111, 112] is more general allowing for more aspects for each elementary layer, but for our purposes we consider each layer as a "usual" possibly weighted and directed network.

Definition 3.1 (Multilayer network)**.** A multilayer network with $L \geq 1$ layers over a set of nodes V is a pair (V_M, E_M), where $V_M = V \times \{1, \ldots, L\}$ and $E_M \subseteq V_M \times V_M$. Each node $i \in V$ is called a *physical nodes*, while $(i, \alpha) \in V_M$ is called *state node* or the *replica* of node i in layer $\alpha \in 1, \ldots, L$. Edges $((i, \alpha), (j, \alpha))$ between replicas in the same layer are called *intra-layer* connections, while edges between replicas $(i, \alpha), (j, \beta)$ with $\alpha \neq \beta$ are called *inter-layer connections*. A multilayer can then be indicated as a quadruplet $M = (V_M, E_M, V, L)$ of nodes (V), state-nodes or node-layer pairs (V_M), edges (E_M) and layers (L). In the special case when inter-layer edges are present only between replica of the same physical node, i.e. for $\alpha \neq \beta$ $((i, \alpha), (j, \beta)) \in E_M$ *iff* $i = j$, the multilayer is called a *multiplex*.

We henceforth indicate by $N = |V|$ the number of physical nodes. Similarly to the matrix representation of single-layer networks, multilayer networks are characterised by tensors.

Definition 3.2 (Adjacency tensor)**.** The multilayer M is completely characterised by its adjacency tensor $\mathcal{M}$, whose components are

$$\mathcal{M}_{j\beta}^{i\alpha} = \begin{cases} 1 & \text{if } ((i, \alpha), (j, \beta)) \in E_M \\ 0 & \text{otherwise.} \end{cases} \tag{3.1}$$

If M is undirected $\mathcal{M}_{j\beta}^{i\alpha} = \mathcal{M}_{i\alpha}^{j\beta}$ and, in general $\mathcal{M}_{j\beta}^{i\alpha} \in \mathbb{R}$ so that the components of the weighted adjacency tensor encode the intensity of interactions between state nodes.

The term $\mathcal{M}_{i\beta}^{i\alpha}$ encodes the intertwining between two replicas; it is a scalar depending, in general, on i, α, β and can also be denoted by $D(i; \alpha, \beta)$ to emphasize the *diagonal coupling*[34, 60, 111] between replicas. From the adjacency tensor we can derive different strength values for each state/physical node: $s_\alpha^{i\alpha} = \sum_{j=1}^{N} M_{j\alpha}^{i\alpha}$ indicates the out-strength of the state node (i, α); summing it over all layers yields the multilayer out-strength of the physical node i without inter-layer connections $s^i = \sum_{\alpha=1}^{L} s_\alpha^{i\alpha} = \sum_\alpha \sum_j M_{j\alpha}^{i\alpha}$, while the out-strength of the state node i in layer α, accounting also for inter-layer edges is $S^{i\alpha} = \sum_j \sum_\beta M_{j\beta}^{i\alpha}$. Observe that the strength of node i

considering only the inter-layer connections of node i is obtained as $S^{i\alpha} - s^{i\alpha}_{\alpha}$. Isolated nodes in traditional single-layer graphs are defined as nodes with degree zero. In a multilayer, conditions on the different types of degree/strength values correspond to different types of isolation:

- $s^{i\alpha}_{\alpha} = 0$—node i is isolated in layer α;
- $S^{i\alpha} - s^{i\alpha}_{\alpha} = 0$—the state node (i, α) has no inter-layer connections;
- $s^{i} = 0$—the physical node i is isolated in the whole multiplex (and we henceforth assume $s^{i} > 0$ otherwise i can be simply removed from V);
- $S^{i\alpha} = 0$—the state node (i, α) is an absorbing state in layer α and when the random walker reaches it remains there with probability 1.

If absorption state are problematic for the wished dynamical model, one can choose[48, 113, 114] to teleport the random walker seated in (i, α) to any (j, β) in the multilayer with uniform probability $\frac{1}{NL}$, which can also be seen as forcing the state node (i, α) to be connected to all state nodes (included itself), and this decreases the occupation probability of the state node (i, α). With these basic definitions we can now extend the concepts of random walks to multilayer networks, and we start from the transition probabilities and the corresponding (random walk) Laplacian operator:

Definition 3.3 (Transition tensor)**.** The probability transitions in one time step constitute the components $\mathcal{T}^{i\alpha}_{j\beta}$ of a rank-4 transition tensor. Different contributions of the multilayer connectivity can be highlighted: $\mathcal{T}^{i\alpha}_{i\alpha}$ is the probability of remaining in the same layer and node, whereas $\mathcal{T}^{i\alpha}_{i\beta}$ with $\alpha \neq \beta$ encodes the probability of *switching* between replicas of the same physical node i in different layers. $\mathcal{T}^{i\alpha}_{j\alpha}$ is the "usual" jumping probability between $i \neq j$ in layer α. Finally, $\mathcal{T}^{i\alpha}_{i\alpha}$ with $i \neq j, \alpha \neq \beta$ is a *switch-and-jump* transition probability.

Definition 3.4 (Normalised Laplacian tensor)**.** The random walk normalised Laplacian tensor is denoted by $\mathcal{L}$ and its components are given by $\mathcal{L}^{i\alpha}_{j\beta} = \delta^{i\alpha}_{j\beta} - \mathcal{T}^{i\alpha}_{i\alpha}$.

For numerical and graphical purposes, see Fig.3.1, we may need to *flatten* the tensors defined above; in this case we apply a flattening function[111, 115, 116] that maps a rank-4 tensor into a $NL \times NL$ supra-matrix, using the same nomenclature of "supra-matrices" introduced in[111].

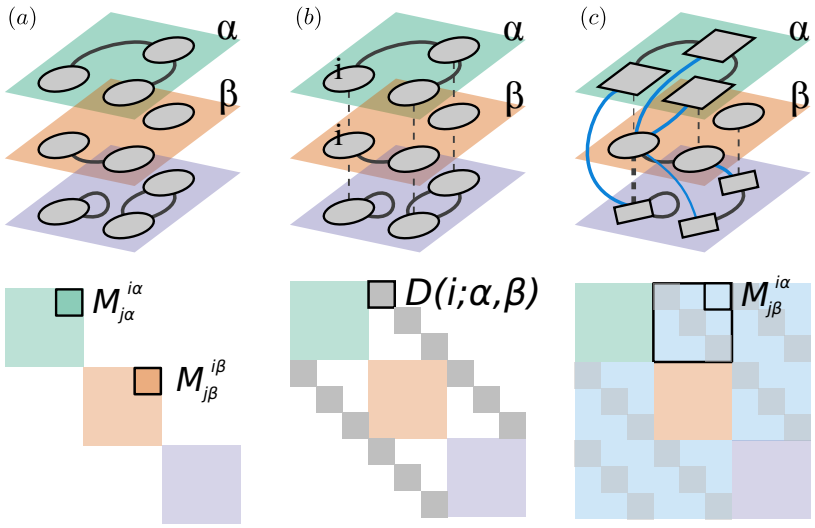

Figure 3.1: Different multilayer networks and their supra-adjacency matrix representation: (a) an edge-coloured multigraph, with layers corresponding to colours and no inter-layer connections; (b) a multiplex network where the replicas in the different layers are interconnected sequentially, a type of intertwining or diagonal coupling, and (c) the most general interconnected case, where inter-layer connections are not restricted to replicas (exogenous interactions). Latin letters denote nodes, while Greek letters are used for layers. $D(i;\alpha\beta)$ is the intensity of the intertwining between state nodes i,α and i,β. $\mathcal{M}_{j\beta}^{i\alpha}$ describes the general inter-layer connectivity. Figure from[7].

3.2 Random walks on multilayer networks

In this new multidimensional set-up, let us first re-write eq. (1.1) for a single-layer network using the tensorial notation[1]. Indicating by δ_j^i the Kronecker delta, by T_j^i the $(i,j)-$th entry of a transition probability tensor and by $\tilde{L}_j^i = \delta_j^i - T_j^i$ the corresponding entry in the random walk normalised Laplacian tensor, then eq. (1.1) reads:

$$\frac{d}{dt}p_j(t) = -\sum_{i=1}^{N} \tilde{L}_j^i p_i(t).$$

Given a transition tensor $\mathcal{T}$ a discrete-time random walk is defined by

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

[1] *Without* the covariant notation and the Einstein summation convention adopted in[112].

and, given its corresponding normalised Laplacian tensor $\mathcal{L}$ we can write the associated continuous-time random walk forward equation

$$\dot{p}_{j\beta}(t) = -\sum_{i,\alpha} \mathcal{L}^{i\alpha}_{j\beta} p_{i\alpha}(t) \tag{3.2}$$

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

If providing the transition rates of a continuous-time Markov chain over a set I completely determines a graph on I, the same does not happen the other way around: given the edge weights of a network we can define not only different continuous-time Markov chains depending on the Poisson process taking place on the nodes or edges, but also different types of discrete-time random walks, as we have seen in Sec.1.2. The additional complexity of the multilayer connectivity translates in an even broader range of random walk dynamics that can be defined on the given multidimensional structure. For instance, a random walker in (i,α) may choose with equal probability adjacent nodes in the same layer and in other layers, or should intra- and inter-layer links be treated differently? This depends both on the real-world diffusion or transport dynamics we choose to model or on the exploration strategy we want to implement for navigating our network, and on the information we have on the system. If there is no information on the set of layers $\{1,\dots,L\}$, in particular on how they are (cor-)related, there is no inter-layer connectivity, and we have the simplest multilayer structure: edge-coloured multigraphs, as shown in Fig.3.1(a).

Definition 3.5 (Edge-coloured multigraphs)**.** An edge-coloured multigraph is a quadruplet $G = (V, E, \{1,\dots,L\}, c)$, where V is the set of physical nodes, E is a multi-set, $\{1,\dots,L\}$ is the set of colours, called layers from now on, and $c : E \to \{1,\dots,L\}$ is a surjective function assigning a colour to each edge.

There are, essentially, two ways to define random walks on these networks. One possibility is to allow the random walker to follow a sequence of edges with different colours: coloured edges are then treated as multiple edges and the degrees of a node counts all the edges, regardless of their colours[117]. This choice is equivalent to performing a classical random walk on the aggregated[112] single-layer network[2]. This

[2]Note that for edge-coloured multigraphs, and only for these, the aggregated and overlay networks are identical

approach is not desirable in general, because one does not know *a priori* if, and to which extent, the information lost through aggregation will affect the results. In other words, is the complex system under study represented as an edge-coloured multigraph or as a weighted network, equivalently? The second approach, adopted in[118], consists in running independent dynamics on each layer and integrating them only afterwards, with the appropriate normalization. The overall transition probability from i to j is obtained as follows:

$$\langle T_j^i \rangle = \sum_{\alpha=1}^{L} m_{i,\alpha} \mathcal{T}_{j\alpha}^{i\alpha}, \qquad \sum_{\alpha=1}^{L} m_{i,\alpha} = 1 \quad \forall i \in V.$$

Observe that the summation is weighted through the factors $m_{i,\alpha} \geq 0$, since each layer can contribute differently, also relative to node i, to the final dynamics of the system. Of course, if $m_{i,\alpha} = \frac{1}{L}$ for all i and α, then $\langle T_j^i \rangle$ is simply the average transition matrix over all layers. If, however, we set $m_{i,\alpha} = \frac{1}{\mu_i} \mathbf{1}_{\{s_\alpha^{i\alpha} \neq 0\}}$, with $s_\alpha^{i\alpha}$ being the out-strength of vertex i in layer α and $\mu_i = \sum_{\alpha=1}^{L} \mathbf{1}_{\{s_\alpha^{i\alpha} \neq 0\}}$ being the number of layers in which i is not isolated, also known as the *multiplicity* of node i, then we can discard the effect of i being isolated in one or more layers and this may be more desirable w.r.t. the first approach, when more information on the diversity of connectivity patterns across layers has to be preserved. Besides, when the edges have both colours and weights only with eq. (3.2) we can guarantee that the scales of weights in the different layers are not accounted appropriately. As a matter of facts, it could be that weights are of ordinal scale on one layer and ratio on another, and a direct summation could lead to errors. In eq. (3.2), instead, for each $i \in V$, we are taking a weighted finite mixture of probability mass functions.

Now, let M be a real possibly directed and weighted multilayer network and let us assume that the edge weights are real numbers encoding the intensity of intra- and inter-layer interactions and that they have comparable scales. In[7] we considered five different random walk dynamics, whose transition probabilities are collected in Tab.3.1.

Classical random walk (CRW) A classical random walker in a state node (i, α) of a multilayer network can move according to intra- or inter-layer connections, with uniform probability. The normalizing factor for the transition probabilities is then the total strength $S^{i\alpha}$.

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

Table 3.1: Transition probabilities for different random walks. (CRW) classical, (PRRW) PageRank, (DRW) diffusive, (MERW) maximal-entropy, and (PrRW) physical with relaxation random walks. $s_{\max}=\max_{i,\alpha}\{S^{i\alpha}\}$; the jumping parameter of the PageRank RW is commonly indicated by α, to avoid ambiguity, we indicate it by r; $\lambda_{\max}$ is the largest eigenvalue of the adjacency tensor and V is its corresponding eigentensor, satisfying $\sum_{i,\alpha}\mathcal{M}_{j\beta}^{i\alpha}V_{i\alpha}=\lambda_{\max}V_{j\beta}$.

PageRank random walk (PRRW) In PageRank[113] a *teleportation* or *jumping* parameter gives the possibility to the walker to reach also nodes that are not directly connected to the current node. In the multilayer PageRank random walk[104] the walker moves with probability r according to its edges (of both types), like in the CRW, and with probability $1-r$ it may *teleport* itself to any state node in the network.

Maximal-entropy random walk (MERW) In the maximal-entropy random walk, as it can be seen in Tab.3.1, the transition probabilities are governed by the largest eigenvalue, and corresponding eigenvector, of the adjacency tensor[112]. Here we assume the network to be undirected so that, as in the single-layer case, we can choose the leading eigenvector to have all non-negative entries (Perron-Frobenius theorem[119]). In the discrete-time framework, it has been shown[119] that all the trajectories of length l between two given nodes have the same probability and that the MERW has maximal entropy production rate among all walks on the network. This characteristic is desirable for processes that aim at a well mixing[120], such as random search over large networks and related localisation of objects or people in real-world systems.

Diffusive random walk (DRW) The jumping process described by the transition probabilities in DRW~Tab.3.1 is characterized by a non-zero probability of remaining

in a given state node (i,β), $1-\frac{S^{i\beta}}{s_{\max}}$, and the probability of following out-going links with a probability normalised over the maximum strength over all nodes, $s_{\max}$. Note that for CRW, PRRW, DRW and MERW, these transition rules generalize the one introduced in[60] for the analysis of multiplex networks.

Physical random walk with relaxation (PrRW) The physical random walk has been defined in[60] to describe those dynamics where the state nodes have a "common memory", so that the information diffuses instantaneously across replicas. Think, for instance, of the system of virtual interactions among individuals, who may have a profile (an alter-ego) in different social networks. A person can then exchange information in a particular social network using the (intra-layer) connections of its alter-ego in that social system, but she/he has always a complete knowledge of the information across the layers. In this case, inter-layer connections between replicas of different physical nodes have no physical meaning and, consequently, are ignored. The physical random walk with relaxation (PrRW)[114] is a variant on the physical random walk, where the assumption on the complete knowledge of intertwining between layers is dropped. It can be seen from Tab.3.1 that its transition probabilities contain a trade-off between intra- and inter-links, which are followed with probability $1-r$ and r respectively. If not differently stated, we consider here $r=0.5$.

3.3 The multilayer diffusion distance

The natural generalisation of the diffusion distance of eq. (1.4) is

$$D_t^2((i,\alpha),(j,\beta)) = \sum_{k,\gamma} \left[p_{k\gamma}(t|(i,\alpha)) - p_{k\gamma}(t|(j,\beta))\right]^2 . \tag{3.3}$$

As for the monoplex case, we can write the transition probabilities in terms of the exponential of the supra-Laplacian matrix, so that eq. (3.3) reads

$$D_t^2((i,\alpha),(j,\beta)) = \sum_{k,\gamma} \left[\left(e^{-t\mathcal{L}}\right)_{k\gamma}^{i\alpha} - \left(e^{-t\mathcal{L}}\right)_{k\gamma}^{j\beta}\right]^2 \tag{3.4}$$

and can be characterized by the spectrum of the Laplacian tensor[112, 115]. At t fixed, $\mathcal{D}_{j\beta}^{i\alpha}(t) = D_t((i,\alpha),(j,\beta))$ is the component of a metric tensor, which can be flattened into a $NL \times NL$ matrix indicated by $\mathbf{D}_t$ and called the supra-distance matrix.

Definition 3.6 (Diffusion distance, tensor, supra-distance matrix, average diffusion distance and normalised supra-distance matrix)**.** Given a multilayer network and a

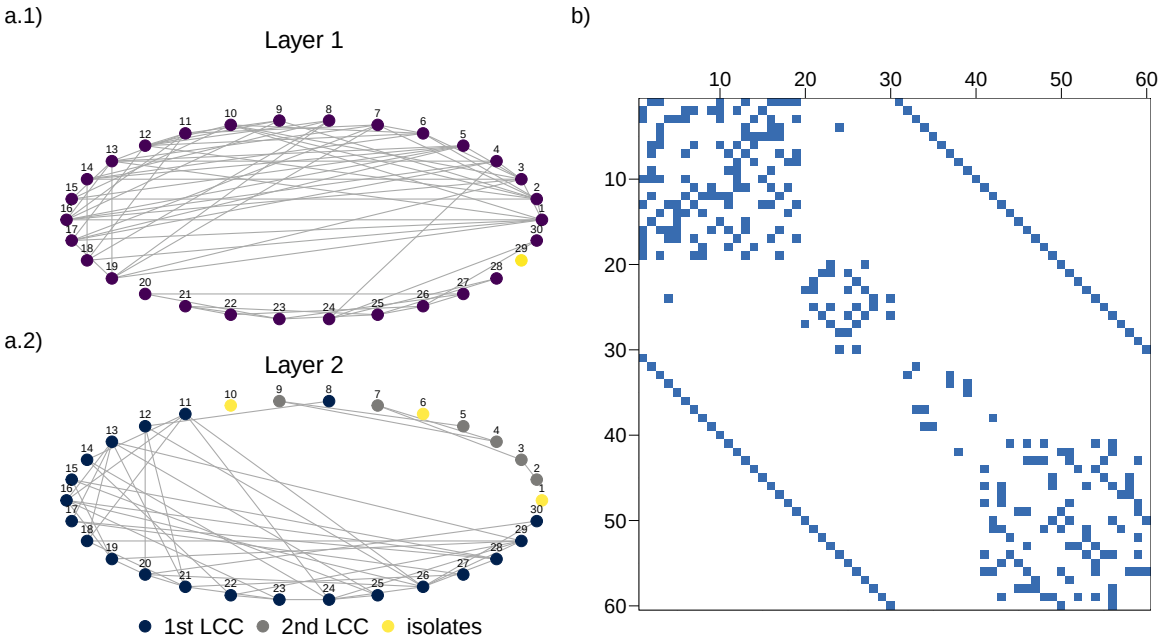

Figure 3.2: A synthetic two-layer multiplex with unitary diagonal coupling $D(i;\alpha,\beta)=1$, for all i and $\alpha\neq\beta$. Panels a) The layers, with isolated nodes, and connected components highlighted in different colours. b) The supra-adjacency matrix of the multiplex.

random walk dynamics on it, whose transition probabilities at time t are given by $p_{j\beta}(t|(i,\alpha)) = \left(e^{-t\tilde{L}}\right)^{i\alpha}_{j\beta}$, we define the diffusion distance between state nodes (i,α) and (j,β) at time t, $D_t((i,\alpha),(j,\beta))$, as in eq. (3.4). The terms of the resulting diffusion distance tensor are denoted by $D^{i\alpha}_{j\beta}(t) = D_t((i,\alpha),(j,\beta))$. We denote by $\mathbf{D}_t$ the diffusion supra-distance matrix, resulting from the flattening of the rank-4 tensor. The average diffusion distance[5] until time $T>0$ is defined as in for the monoplex

$$\bar{D}_T = \frac{1}{T}\sum_{i=1}^{n} D_{t_i} \qquad \text{for } 0<t_1<\cdots<t_n\leq T$$

The corresponding average diffusion supra-distance matrix is indicated by $\bar{\mathbf{D}}_t$. Finally, we introduce the normalised diffusion supra-distance matrix[5]

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

Figure 3.2 shows the two layers—panels a)— composing a two-layer multiplex with diagonal coupling, $D(i;1,2)=D(i;2,1)=1$, and its supra-adjacency matrix—panel b). In each layer, we have a network with $N=30$ nodes generated from a stochastic block model[121, 122] with two blocks. We chose the probabilities in order to have a diverse topology: dense groups, disconnected components and isolated nodes.

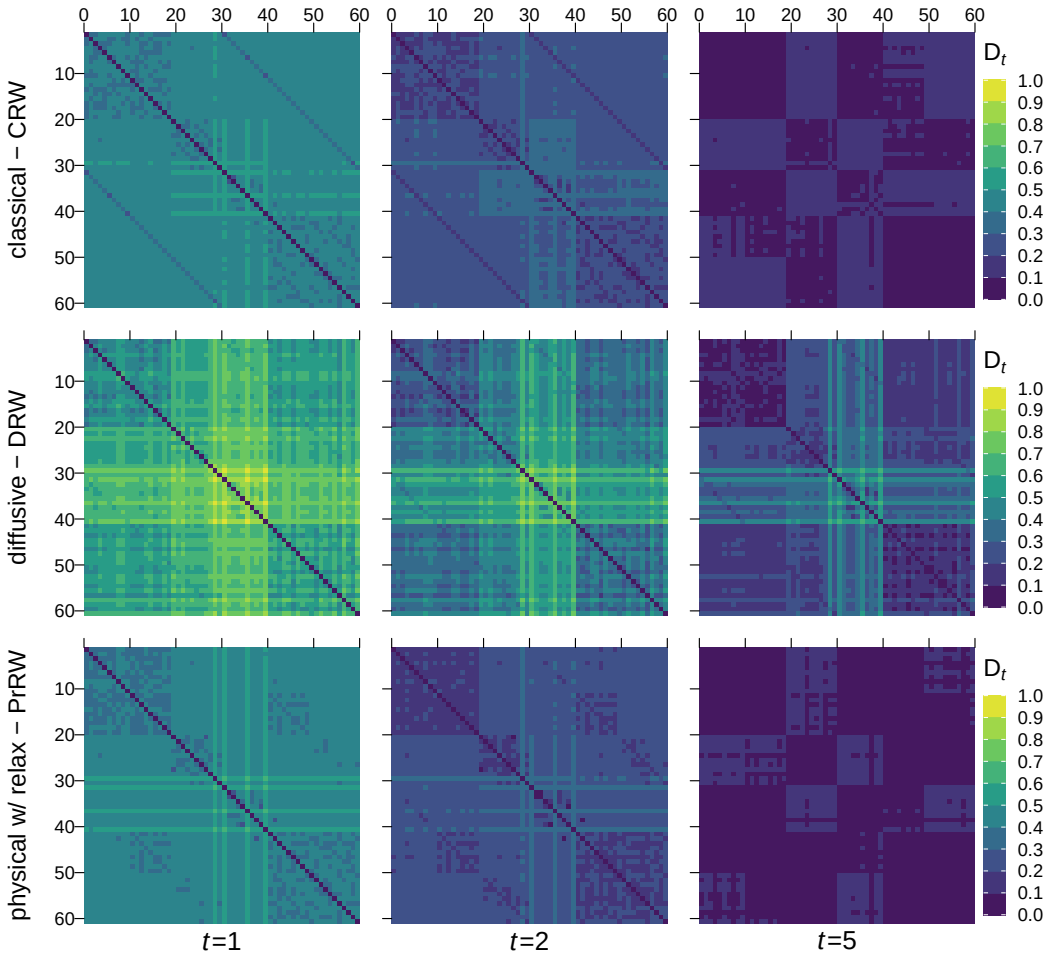

Figure 3.3: Diffusion distances for the synthetic two-layer network of Fig. 3.2 functional characterization. The $NL \times NL$ diffusion supra-distance matrices for three different random walk dynamics and diffusion times $t = 1, 2, 5$. The PrRW is evaluated for $r = 0.5$. Distances have been rescaled in $[0, 1]$ dividing them by $\sqrt{2}$. Distances w.r.t. the DRW are in general larger and isolated or weakly connected nodes lie far away from the other nodes.

A random walker exploring one of these layers only, would be trapped in the connected component where the walk starts, instead, due to the inter-layer links the multiplex becomes connected, and the random walker can move across different components switching between layers. Figure 3.3 shows the supra-distance matrix $\mathbf{D}_t$ for the three more diverse RW dynamics (rows) on a synthetic multilayer network, for different values of the diffusion time t (columns). State nodes in supra-matrices are numbered from 1 to NL. Recall that the diffusion time plays the role of a scale parameter[5] and that the continuous-time Markov chain has exponentially distributed holding times with rate $\lambda = 1$, i.e., the expected time occurring among each step of the RW is 1. $D_{t=1}$ is then a function of the micro-scale structure of the multiplex, and here the isolated nodes are clearly visible. Here, distances have been rescaled in

$[0, 1]$ dividing them by $\sqrt{2}$ and the colour bar has been set accordingly. Remarkably, the distances w.r.t. the DRW span a larger interval than the others and isolated or weakly connected state nodes are more distant from the rest of the units. This is a consequence of the diffusive dynamics, where the jumping probabilities do not depend on the vertex, but on the strength of hubs, so that a random walker leaves a hub much faster than a low-degree node. It is also worth noticing that, those nodes, which are disconnected from the largest connected component of the second layer (nodes 1, 6 and 10), are generally nearer to the nodes in the first layer than to the nodes in their same layer (dark upper and lower diagonals).

To unveil the persistent meso-scale structures of the multiplex, we average the diffusion supra-distance matrices for up to $t_{max} = N$[5] and run a hierarchical clustering on the average diffusion supra-distance matrices, summarizing the results in Fig. 3.4. As expected, the presence of inter-links moves loosely connected nodes closer to their replica in the other layer, and the clustering does not separate the two layers. In this and the following plots, the state nodes in the supra-distance matrices are re-ordered according to the dendrogram returned by the complete linkage method for hierarchical clustering, (R) **hclust**[123].

At the beginning, the two nearest nodes are grouped together, then at each step, the distance between the clusters is defined as the maximum distance between their nodes and the two nearest clusters are aggregated, until all the nodes are in the same group. In[5] it has been shown that the dichotomy between fast-shrinking (intra-community) and slow-shrinking (inter-communities) distances probes the meso-scale structure at different resolutions and that the most persistent and representative meso-scale structure is the one maximizing the average diffusion distance between clusters. Here, we use the dendrograms just as a guide for the eye and do not delve into the multilayer community detection[114, 124] or the analysis of the hierarchies[125].

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

$$D_t^{\mathrm{eq}}(i,j) = \left(\sum_{\alpha=1}^{L} \frac{1}{D_t((i,\alpha),(j,\alpha))} \right)^{-1}. \tag{3.5}$$

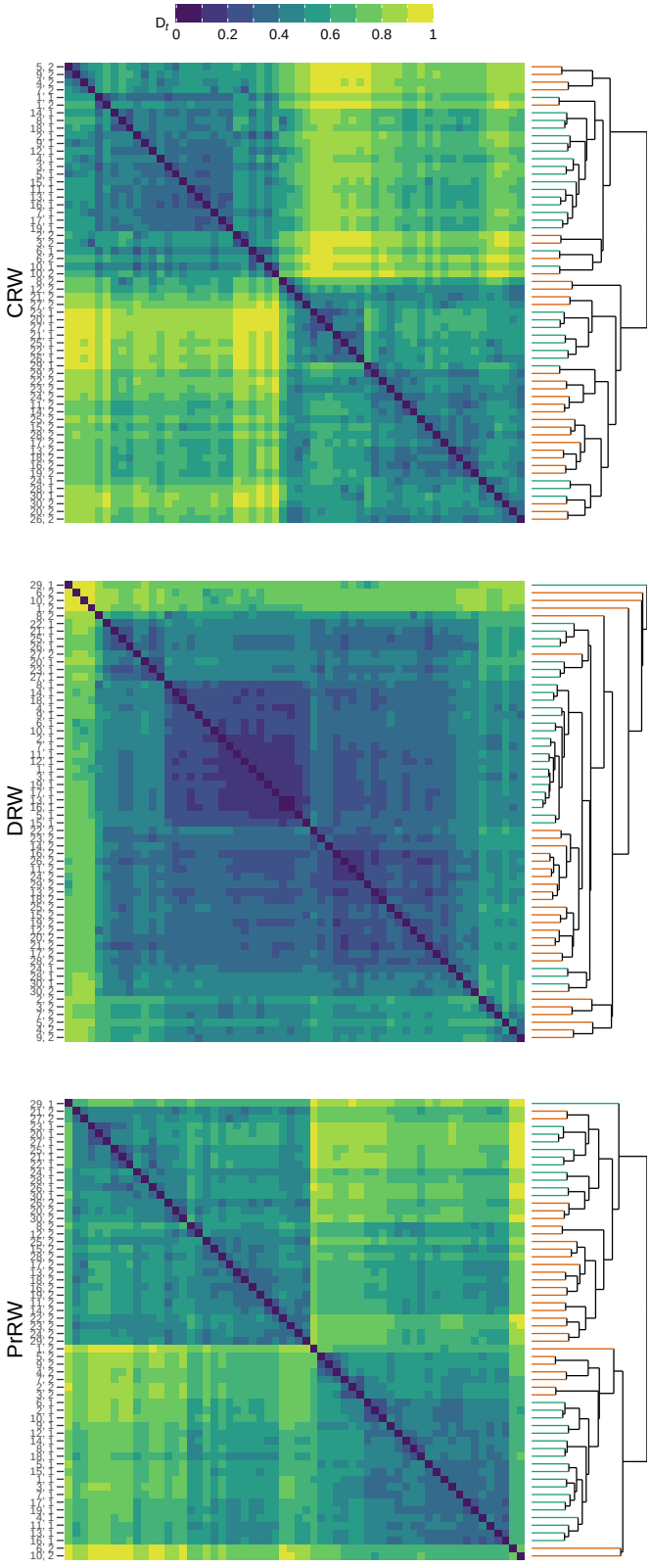

Figure 3.4: Average diffusion supra-distance matrices of the two-layer synthetic network shown in Fig. 3.2, w.r.t. three RW dynamics: classical (CRW), diffusive (DRW), and physical with relaxation $r = 0.5$ (PrRW) as in Fig. 3.3. The state nodes (i, α) are coloured in the corresponding dendrogram according to the layer they belong to (green for layer 1 and orange for layer 2). The CRW supra-distance matrix has two main blocks, which do not perfectly correspond to the layers, as we can see that nodes that are not in the LCC of layer two tend to be closer to nodes in layer 1. The DRW preserves better the layer distinction, with isolated and weakly connected state nodes being isolated also in the clustering (e.g. yellow block in the top left corner). Also in the clustering w.r.t. the PrRW diffusion distances nodes from the same layer tend to be aggregated earlier in the hierarchy with isolated state node clearly visible.

Note the equivalent diffusion distance (briefly, *equivalent distance*, if not ambiguous)

is as an aggregation of diffusion distances across layers, and it is not related to the concept of resistance distance[55].

Another way to obtain the mutual distances between physical nodes, consists in aggregating the multilayer network[112] and then compute the diffusion distances on the aggregated single-layer network. Observe, however, that the equivalent distance matrix, denoted by $\mathbf{D}_t^{\text{eq}}$, is quite different from the diffusion distance matrix on the aggregated network, as shown in Fig. 3.5.

Looking back at the network plot, Fig. 3.2, we can see that the nodes from 1 to 19 are densely connected in layer 1, while node 8 and nodes from 11 to 30 form the largest connected component of the second layer. Both, the equivalent and aggregated distance matrices show a clear block corresponding to the last ten nodes. However, only the equivalent diffusion distance matrix captures the particular position of the first ten nodes (removing 8) in the cluster hierarchy: in the multiplex they are distant from the last ten, because they belong to different communities in layer 1 and to disconnected components in layer 2. We can also discard the inter-layer edges and consider the resulting edge-coloured multigraph, computing the transition probabilities as in (3.2) and, only then, compute the diffusion distances. Figure 3.6 shows this distance matrix w.r.t the CRW and averaged over time. Distances in the edge-coloured multigraph are similar to the distances in the aggregated network, with only minor permutations inside the clusters.

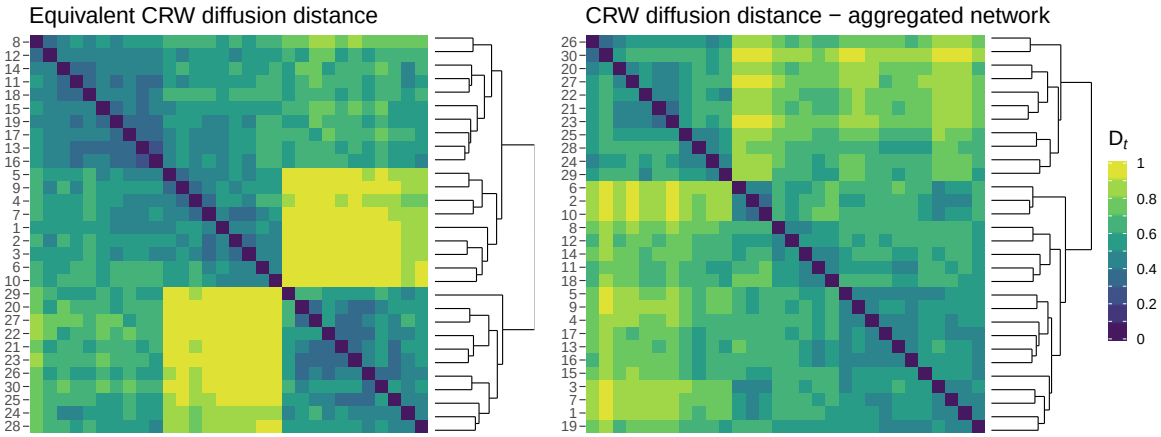

Figure 3.5: Comparing different averaged diffusion distances between physical nodes, regardless of layers. The equivalent diffusion distance matrix obtained with the reduction of Eq. (3.5) and the diffusion distance matrix w.r.t. the CRW and averaged over time of the aggregated network corresponding to the synthetic multiplex of Fig. 3.3. The distances have been rescaled to $[0, 1]$ dividing by the corresponding maximum over all node pairs (equivalent distances: $\max(\bar{\mathbf{D}}_t^{\text{eq}}) \approx 0.11$, aggregated: $\max(\bar{\mathbf{D}}_t) \approx 0.13$).

As in the single-layer case, through random walk dynamics we can map state

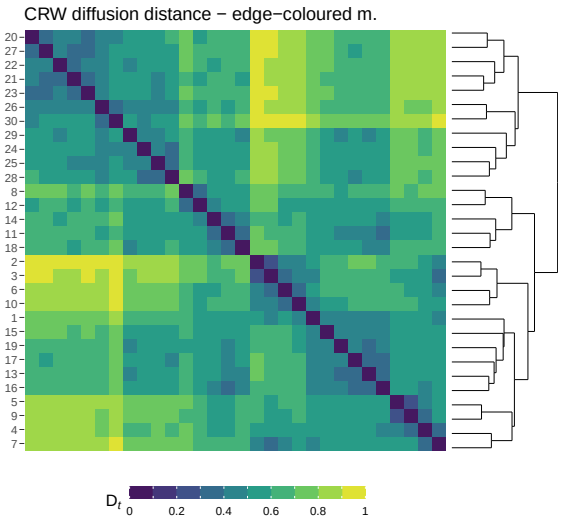

Figure 3.6: The average diffusion distance matrix of edge-coloured multigraph obtained discarding the inter-layer edges. The random walk is the CRW and the transition matrix of the multigraph is obtained as in (3.2). Here we used $m_{i,\alpha} = \frac{1}{\mu_i}\mathbf{1}_{\{k_\alpha^{i\alpha}\neq 0\}}$, with $k_\alpha^{i\alpha}$ being the degree of vertex i in layer $\alpha \in 1, 2$ and μ_i the multiplicity of node i, so that we prevent the random walker being trapped into the isolated nodes of each layer. The distance matrix has been rescaled to $[0, 1]$ dividing by the maximum over entries, as in Fig. 3.5 (edge-coloured: $\max(\bar{\mathbf{D}}_t) \approx 0.16$).

nodes to points in space: indeed $p_{j\beta}(t) = \tilde{p}_l(t)$, with $l = j + \beta$ can also be seen as the component of a supra-vector in R^{NL}. We then generalize Def. 1.2 as follows:

Definition 3.7 (Diffusion space - 2)**.** The set of state nodes $\{(i, \alpha) : i = 1, \ldots, N,\ \alpha = 1, \ldots, L\}$ endowed with the diffusion distance at a fixed value of t, is a metric space. A random walk dynamics maps each state node (i, α) to a point in the space $\mathbb{R}^{NL}$. Furthermore, as the Euclidean norm of the probability supra-vectors, with components $p_{j\beta}(t|(i, \alpha))$, is bounded, the state nodes are embedded into a bounded subspace of $\mathbb{R}^{NL}$, which is here referred to as the *diffusion space* or *diffusion manifold* or *diffusion embedding*.

We can now study more realistic synthetic and real multilayer networks through the lens of diffusion geometry. We are particularly interested on the effect of layer-layer correlations and of the intra-layer connectivity on different diffusion dynamics. In the case of undirected unweighted multiplexes it has been shown[34, 115] that the strength of the diagonal coupling can either hinder—when low—or speed up—when high—the convergence of a classical random walk to its stationary distribution. This has implications, for example, in city transportation networks: the addition of new transport media (new layers) alone may not be sufficient to support human flows, changing from bus to train, e.g., should also be easy and affordable (strong coupling between replicas). But what if we cannot increase the coupling strength between replicas? Could additional links between replicas of different physical nodes—hence moving from a multiplex to a more general inter-layer connectivity—help diffusion in the whole system? In terms of our transportation network example, a diagonal coupling means that a person can move from the train transportation layer to the bus in the same station/bus stop. A bottleneck can emerge when too much people take

the same bus in that train station, but if other bus stops are easily reachable by walk from the train station, passengers can spread avoiding queue congestions. In the next section, we will investigate the effect of inter-layer correlations and connections on the diffusion spaces.

3.4 Multilayer Diffusion Manifolds

The use of different random walk dynamics to explore a system has an impact on the distances between its units and, consequently, on how the units are distributed in the induced diffusion spaces. Similarly, the diffusion time shapes the pairwise distances, highlighting local features of complex network geometry on short time scales and its more persistent structures for large diffusion times. In the multilayer setting, there is an additional level of complexity given by the inter-layer connections and by the layer-layer correlations.

To gain further insights, we generate 3 distinct classes of synthetic multiplexes, with system size $N = 200$, and analyse them through the lens of diffusion geometry. The first class consists of Barabási–Albert (BA) scale-free networks[102] on each layer: we consider a linear preferential attachment with 4 edges added by each new node during the growth process, while setting at 10% the edge overlapping across layers—defined in terms of the fraction of links which are present in both layers among the same pairs of nodes[104]. The second class consists of Watts–Strogatz (WS) small-world network[21] on each layer, obtained by rewiring lattices with probability 0.2, where edge overlapping is tuned similarly to the first class. The third class consists of layers with strong meso-scale structure organized in 4 groups, like in a Girvan–Newman (GN) model[62] on each layer: the probability that two nodes within the same group are connected is 1, whereas cross-group connections are much sparser and exist with probability 0.05. Group overlapping[104]—defined in terms of the fraction of nodes planted in the same group on both layers—is fixed at 1%.

The role of layer-layer correlations and their interplay with the distinct network topologies considered above is summarized in Fig. 3.7. As expected, there are relevant differences due to the type of random search dynamics and to the topological features of the underlying topologies. For instance, the diffusive walk for the BA system leads to a high level of mixed pathways across layers: in the resulting diffusion embedding the nodes are not grouped together according to their layer assignment, as can be seen from the mix of green and purple in the corresponding dendrogram[3]. For the same

[3]The distance matrices and corresponding dendrograms are small, but here we are interested in

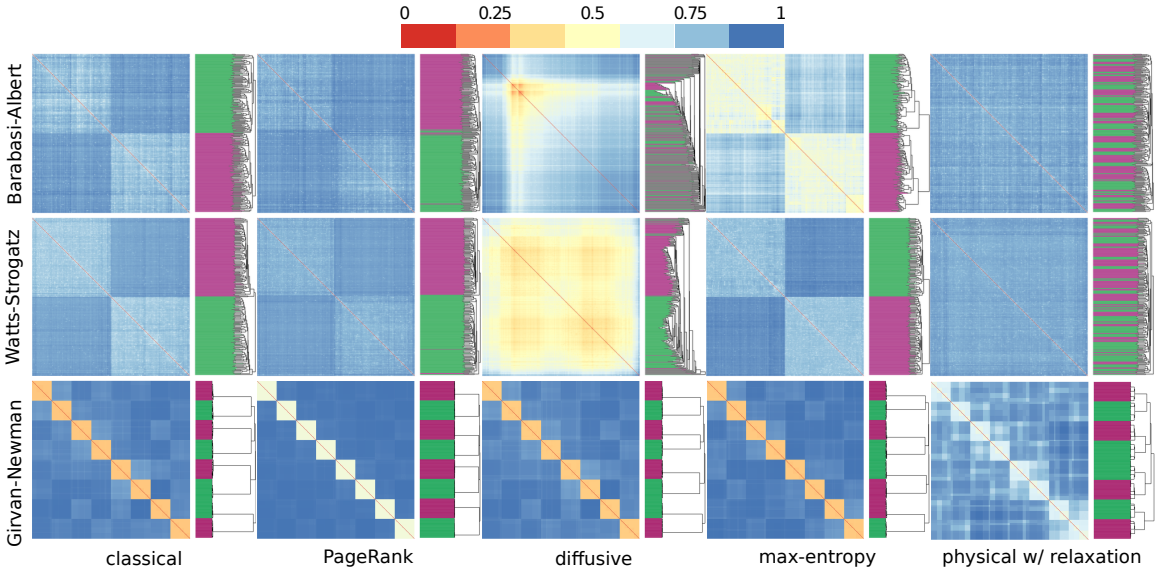

Figure 3.7: Average diffusion distance $\bar{D}_t$ on two-layer multiplexes with different topologies, for fixed values of global average edge overlap (BA and WS) and partition overlap (GN) between layers. Dendrograms on the right-hand side of each distance matrix represents the corresponding hierarchical clustering to highlight the meso-scale organization of the system, with colour encoding the planted node assignment in each layer. Distances have been rescaled dividing by the respective maximum $\frac{\bar{D}_t}{\max \bar{D}_t}$. The BA model is characterized by the presence of hubs, which are clearly recognizable in the supra-distance matrix w.r.t. the diffusive random walk, despite the small overlap between layers. The GN two-layer multiplex has a meso-scale organised in strong communities, partly overlapping across layers. Note that at variance with edge overlapping, here partition overlapping is defined in terms of nodes belonging to the same group without requiring those nodes to be connected by an overlapping edge. Figure from[7].

walk, in the case of the WS network the result is the opposite: nodes aggregate into functional clusters that are highly homogeneous w.r.t. the nodes layer assignment. It is also not surprising that a high amount of *geometric mixing* (i.e. the mixing in the embedding space of state nodes that belong to different layers or communities in the network) is observed, on all models but the GN, when the network functionality is studied through the physical random walk with relaxation. Instead, this dynamics on the GN model makes intra- and inter-layer distances more homogeneous than the other RWs and the cross-layer communities appear only for very large values of partition overlap, see Fig. 3.8. Overall, it is not guaranteed that diffusion pathways across layers favour the geometric mixing in the diffusion manifold: the result depends on the type of dynamics and on layer-layer correlations.

"the big picture".

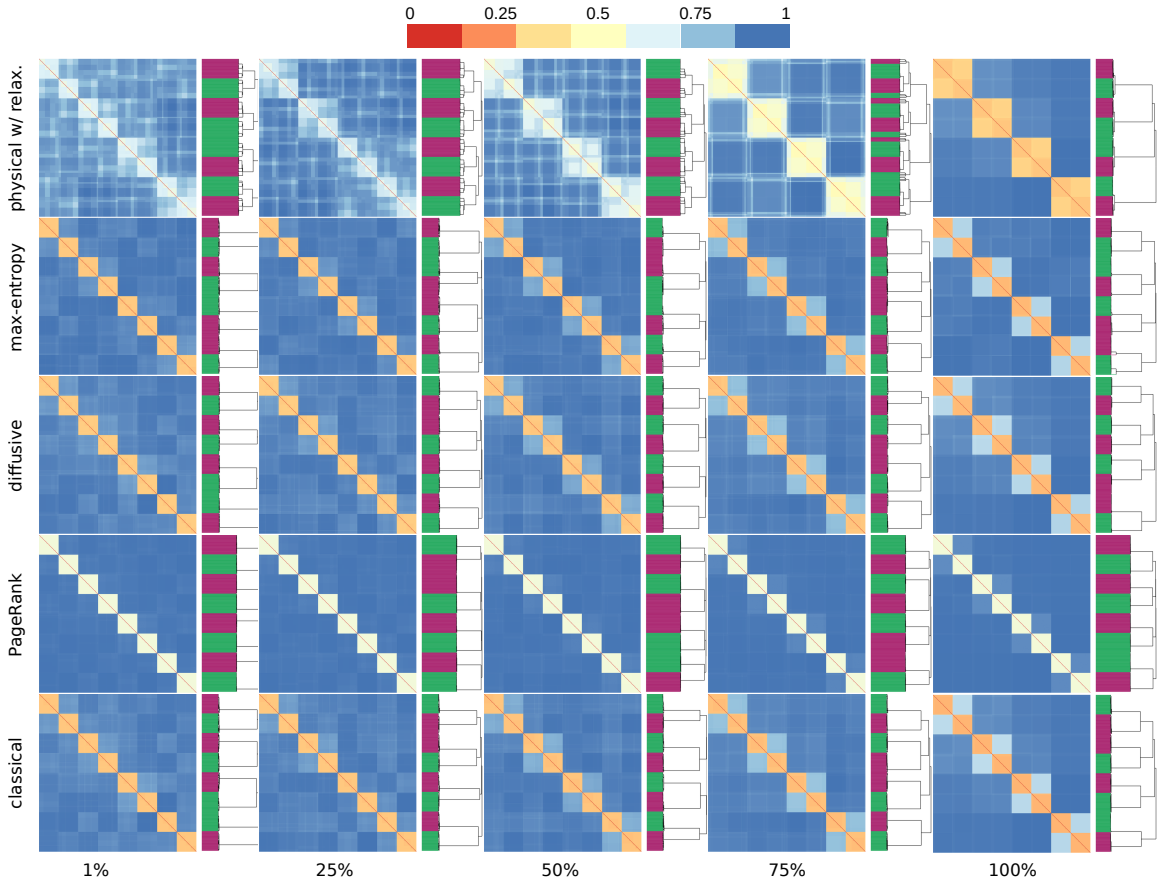

Figure 3.8: Average diffusion distance $\bar{D}_t$ for a two-layer multiplex with meso-scale organization in strong communities (GN model as in Fig. 3.7), partly overlapping across layers (percentage on the x-axis). Note that at variance with edge overlapping, here partition overlapping is defined in terms of nodes belonging to the same group without requiring those nodes to be connected by an overlapping edge. As the overlap between the communities in the two layers grows, macro-communities across layers emerge, more—PrRW—or less—PRRW—clearly. Figure from[7].

To further explore this interplay between structure and dynamics, we have considered a second battery of synthetic models, where we increasingly add inter-layer connectivity between layers. The absence of information pathways across layers, happening for instance when two layers are not coupled together, leads naturally to disjoint diffusion manifolds, i.e. disjoint clouds of points in the diffusion space, each one corresponding to the distinct layers. When the two layers are interconnected together, a trivial result is that the strength of inter-layer connectivity facilitates the flow of information across layers. However, the above process hides an interesting phenomenon, that is unveiled in Figs. 3.9-3.10. To better characterize it, we calculate

the Frobenius norm, which is defined as follows, for a generic matrix $\mathbf{A}$

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

to quantify the overall intensity of an average diffusion distance matrix. Observe that $\|\mathbf{D}_t\|_{\mathrm{F}} = \sqrt{2}N\sqrt{\mathcal{D}_t}$, where $\mathcal{D}_t$ is the ASDD (average of the squared diffusion distances, eq. (1.6)), a measure of spread, so that also the Frobenius norm of the supra-distance matrices quantify the dispersion of nodes in the diffusion space. The Frobenius norm is computed when the two layers are not coupled, on the union of the two distance matrices, i.e., $\sqrt{\|\mathbf{A}\|_{\mathrm{F}}^2 + \|\mathbf{B}\|_{\mathrm{F}}^2}$, and then on the supra-distance matrices for increasing fraction of inter-layer connectivity: first state nodes corresponding to the same physical node are interconnected with each other to create an interconnected multiplex; after that this regime is reached, the cross-links between state nodes corresponding to all other physical nodes are created, until the total of N^2 connections is generated. Remarkably, when one interlink is added between the layers, the Frobenius norm increases: this is due to the fact that the new link coupling the two layers creates a bottleneck for the information exchange across layers, even for large values of t. Once that more inter-links are added, the Frobenius norm decreases, tending to a plateau when the fraction of inter-layer links approaches the 100%. It is worth noticing how the structures of the single layers and of the whole fully-interconnected multilayer coexist in the different diffusion spaces w.r.t. the diffusion time t. In case of BA networks in each monoplex, Fig. 3.9, their typical heterogeneity (presence of hubs) characterizes also the diffusion manifolds. As a matter of fact, hubs have smaller diffusion distances to any other node. On the contrary, for small values of t we can see in Fig. 3.10 that the community structure is concealed by the presence of the inter-layer links. This is totally expected, since diffusion time plays the role of a scale parameter and locally, nodes in the fully-interconnected multilayer with communities, Fig. 3.10-(d), are very similar.

Our results highlight that the existence of topological correlations across layers induce changes in how information is exchanged between state nodes. Such changes alter diffusion distances and might lead to two different regimes: i) flow keeps segregated within layers and the diffusion manifold corresponding to the multilayer consists of two well separated sub-manifolds representing each layer separately; ii) flow is integrated and the diffusion pathways across layers mix up those sub-manifolds.

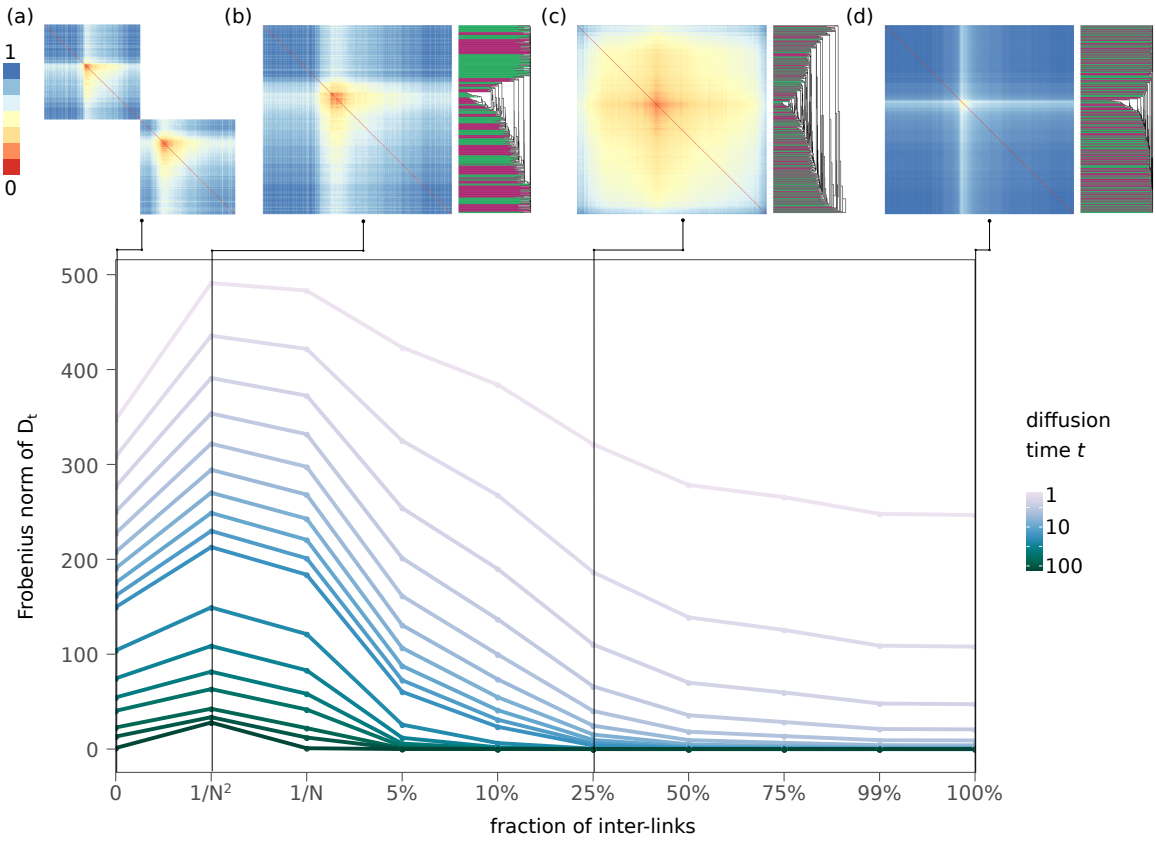

Figure 3.9: Frobenius norm of the supra-distance matrices of a synthetic two-layer network, w.r.t. a diffusive random walk. Each layer is generated from a BA model as in Fig. 3.7. As the fraction of inter-layer links grows, we move from two disconnected multiplexes to a fully inter-connected multilayer with all N^2 connections across layers. The heatmaps of $D_{t=5}$ are four representatives of the different regimes: (a) uncoupled layers; (b) a single-inter-link between the replicas of a random node coupling the two layers, i.e., $\left[\frac{1}{N^2}, \frac{1}{N}\right]$ partially interconnected multiplex; [1/N] fully interconnected multiplex (all state nodes corresponding to the same physical node are interconnected); (c)-(d) $\left[\frac{1}{N}, 1\right]$ multilayer regime consisting of an interconnected multiplex with the addition of cross-links between state nodes of distinct physical nodes. Figure from[7].

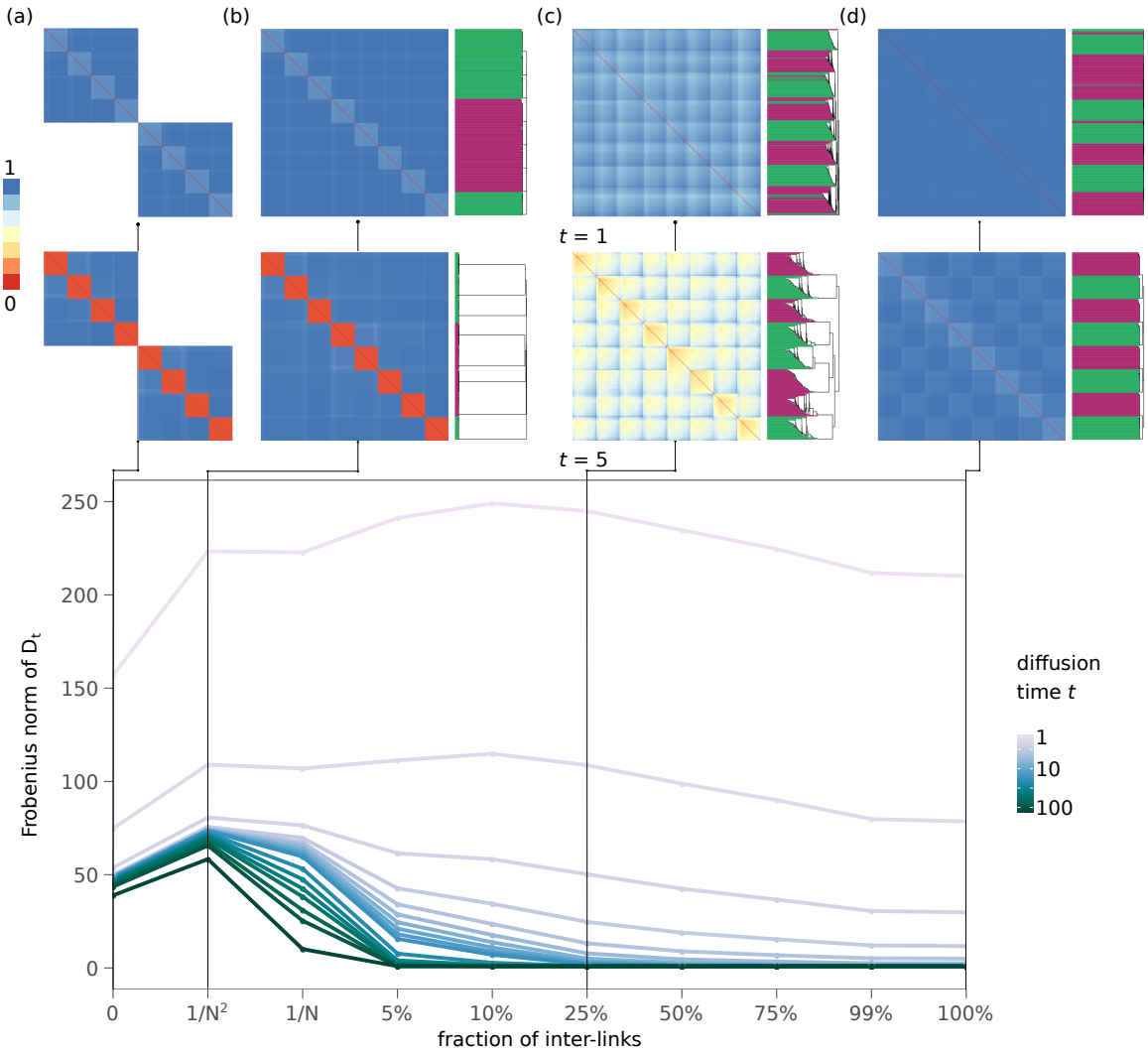

Figure 3.10: As in Fig. 3.9 for networks with a meso-scale organized in four strong communities in each layer and partition overlap of 1%. The heatmaps on the top correspond to $D_{t=1}$, while the other are $D_{t=5}$. Figure from[7].

Real multilayer networks

We use the newly introduced family of metrics to study two real systems with multiple types of interactions: the multimodal transportation network of London[60] and the multilayer Noordin Top terrorists network[126]. The first system consists of three layers corresponding to the Tube, overground, and DLR, arranged in a multiplex with couplings $D(i;\alpha,\beta) = 1$ for $\alpha \neq \beta \in \{1,2,3\}$. Nodes represent stations ($N = 369$ in total) and connections between them are weighted and undirected. The network of interactions among 78 Indonesian terrorists ($N = 79$ in the data set, but actor 58 is usually removed since it is disconnected in all layers) is a four-layers multiplex, representing their pairwise trust (T), operational (O), communication (C) ties, and business (B) relations[127].

Figure 3.11 shows the diffusion manifold corresponding to $\bar{D}_t$ (the average diffusion manifolds), projected in $\mathbb{R}^3$ through multidimensional scaling[128] (see Appx. A.3), induced by different RW dynamics. As observed in[60], the best exploration strategy, i.e. the RW to prefer for covering efficiently the network, depends on the topology of the multilayer. This is reflected in the maps of Fig. 3.11, even though they are low-dimensional approximations of the true diffusion manifolds. As a matter of fact, for the London transportation network, the manifolds induced by the classical, PageRank, and diffusive random walks appear qualitatively very similar with each other, and considerably different from those induced by MERW and PrRW. Instead, the supra-distance matrices and manifolds obtained for the terrorists network appear similar in that all have a group of nodes with small pairwise distances, and another group of nodes which are distant from each other. To quantify more adequately how diverse, or similar, the manifolds are, we compare their supra-distance matrices by means of Mantel's test[129, 130], where the null hypothesis is that the pairwise distances in one matrix are not monotonically related to the corresponding distances in the second matrix, and show the results of our test in Figs. 3.12-3.13.

Discussion and conclusion

In this chapter we have considered different random walk dynamics on multilayer networks, from the classical one to various biased random walks, which either extend existing dynamics on single-layer networks or are defined ad-hoc for the multidimensional setting. The multilayer diffusion geometry here introduced, as for the single-layer diffusion geometry, provides a framework for analysing the functional shape of complex multilayer networks, through network driven diffusion processes,

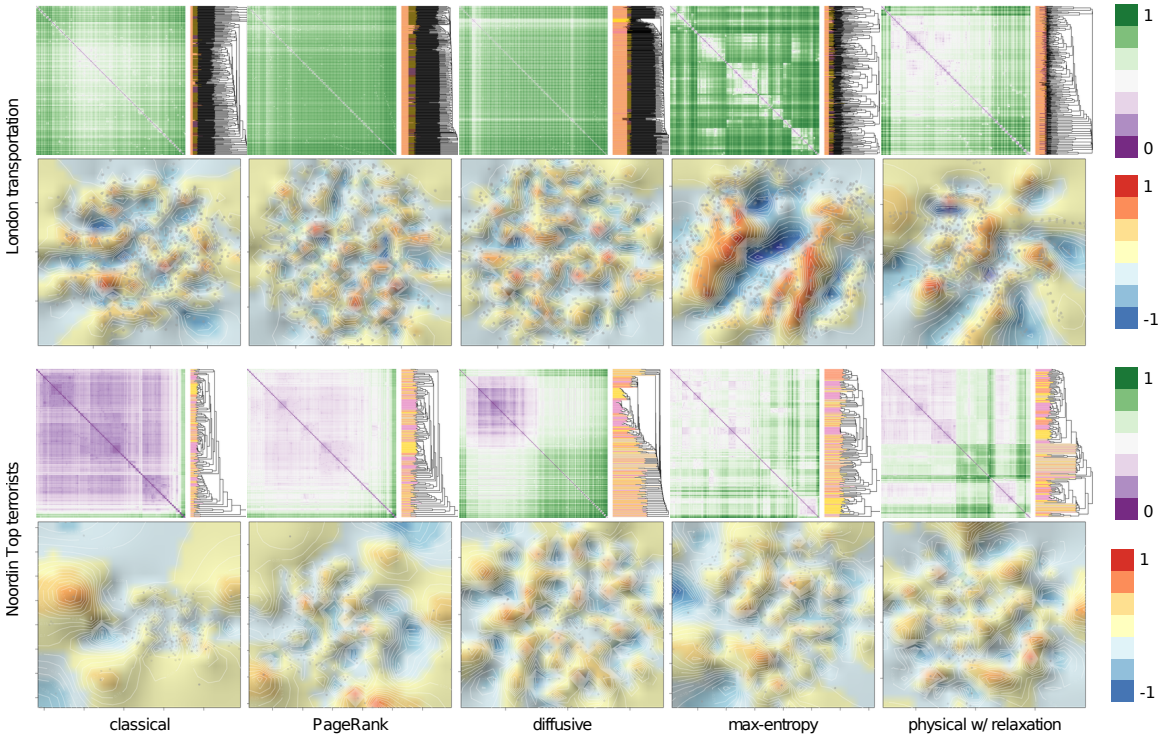

Figure 3.11: Average diffusion distance supra-matrices $\bar{D}_t$ organized according to their hierarchical clustering as in previous analyses, and projection in $\mathbb{R}^3$ of the corresponding diffusion manifolds, in the case of the two real multiplex networks: the London public transportation network (top panels) and the social relationships of Indonesian terrorists (bottom panels). In both cases, nodes have been embedded in space, more specifically in $[-1, 1]^3$, through multidimensional scaling, a metric preserving embedding method that depends on two parameters: a distance (or dissimilarity) matrix and an embedding dimension[131] (see Appx.~A.3 for details). To facilitate the visualization of the three-dimensional embeddings, we draw the surface which better approximates the cloud of nodes in $\mathbb{R}^3$ and project it on the plane, encoding the third dimension with colours. Nodes are shown as dots on the top of the surfaces. Figure from[7].

which also approximate metastable synchronization, consensus, and random search dynamics.

With the analysis of synthetic multilayers, we have provided evidence that the multilayer diffusion distance, and the induced manifolds, encode information due to both the structure of the multilayer and the type of random walk dynamics on it, highlighting that the two—structure and dynamics—cannot be separated when inferring information about the function of a system. Of course, here we focus only on averaging operators (Laplacians) and diffusion processes for studying the function of networks, while other dynamics may display different functional behaviours[132]. In the same spirit, different representations of complex systems accounting for other types of interactions, e.g. hypergraphs and simplicial complexes, may provide diverse

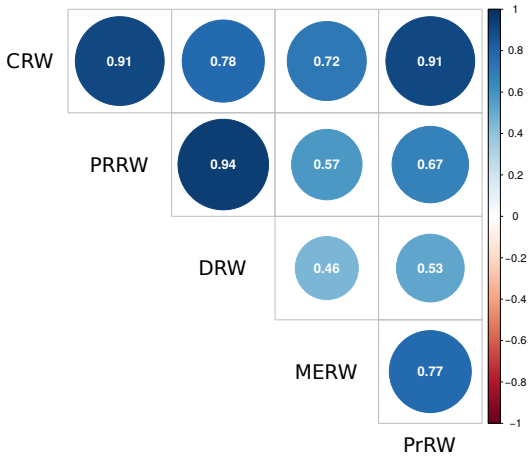

Figure 3.12: Estimating the similarity between diffusion manifolds corresponding to different RW dynamics, evaluated on the London multimodal transportation network. We calculate the Pearson's correlation (encoded by size and colour) between the entries of pairs of supra-distance matrices (Mantel's statistic) which is then tested for significance by permutation (permutation test), with $\alpha = 0.001$. The test is general, because it applies directly on distance matrices, whereas any test performed on the low-dimensional embedding of diffusion manifolds would be less precise because of the information loss during projections.

insights on the functionality of complex systems[51, 133].

From the analysis of synthetic networks with overlapping edges or groups across layers, we have found that the interplay between dynamics and topology cannot be easily decoupled: e.g., the classical random walk reveals a manifold in which state nodes of the different layers have larger distances than the intra-layer distances, but this does not remain true in the presence of strong communities, which are not necessarily overlapping. Also the behaviour of the MERW is not trivial: the top-level hierarchical structure unveiled in the Barábasi–Albert model is compatible with that of the Watts–Strogatz network, despite the high heterogeneity of the first, and this could be surprising since MERW is influenced by irregularities in nodes degree. In another scenario, where two layers are originally uncoupled and do not exchange information, we add inter-layer connectivity to better understand how the originally disjoint diffusion manifolds approach each other because of the presence of multilayer

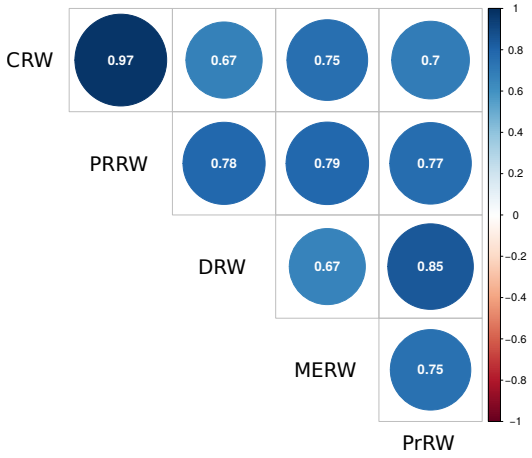

Figure 3.13: Same as in Fig. 3.12 for the Noordin Top terrorists network.

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

Finally, we have applied our novel framework to two empirical multilayer systems, namely the public transportation of London and the social network of Noordin terrorists. The diffusion geometry corresponding to different random walk dynamics are not necessarily distinct, and we have developed a quantitative method to assess the correlation between the underlying multilayer diffusion manifolds. In the case of the transportation system, we find that the MERW, which in the synthetic networks was able to separate the layers, highlights two groups of near nodes, that are not captured by other dynamics. This may suggest that (i) the structure of this system, at different scales, has features that are characteristic of different models (e.g., heterogeneity and communities); (ii) different dynamics induce different manifolds and consequently, the analysis of networked systems embedded into space cannot exclude the analysis of the dynamics itself. Conversely, in the case of the social system, we find that the metrics have generally higher correlations, so that their latent diffusion spaces may be likewise similar. The hierarchical structure unveiled by the supra-distance matrices seems to suggest a cross-layer core-periphery functional organization.

In this work we extended the mathematical framework for diffusion distances, introduced in[5], to different random walk dynamics and different interconnected layers. We highlighted some interesting results in the sense of functional meso-scale structures. Nonetheless, the thorough analysis of functional multilayer communities is beyond the scope of this work and left for future work. Our work provides a novel tool for the analysis of multilayer systems from a network geometry perspective[4]. Since the latent diffusion geometry is induced by network-driven processes, our framework provides also a complementary view to structural analysis, such as the one provided by hyperbolic network geometry[3, 134, 135], recently used to analyse multilayer networks[136], and higher-order analysis[137].

Chapter 4

The functional rich-club

Real systems are characterized by complex patterns of interactions between their units, by dynamical processes evolving over the units and their connections, and by the interplay of the two. In the previous chapters we have seen how particular micro-, meso- and macro-scale structures affect the dynamical patterns not only locally but also globally. For instance, dynamics evolve differently on networks characterised by different degree or strength distributions and this in turn affects other features of the dynamics, as its global efficiency in case of parallel communication.

The degree of a node and the length of shortest paths between node pairs are among the most used and useful statistics of the topology of a network, and are at the basis of more complex descriptors characterising the networks at different scales. A particular structure involving a group of nodes (hence in the meso-scale) but which needs information on the whole network (macro-scale) to be defined, is the rich-club. The rich-club phenomenon is defined in terms of an excess connectivity among nodes with high degree (also called rich nodes or hubs) that cannot be explained by pure chance[138, 139]; it is the tendency of rich nodes to establish denser than expected and sometimes costly interlinks among them. In connectomes, networks representing the structural connections in the brain, it has been shown that hubs play a key role in the global exchange and processing of information, and form a rich club that enables the integration of the information processed by physically segregated modules[140–145]. In a recent work[146], Moretti and Hütt studied the role played by hubs, cores and rich-clubs in self-organized wave and excitation patterns, finding that these act as organizers and facilitators for the emergence of the global, collective patterns both in network models and in the human connectome. In network neuroscience, it has been suggested that the rich-club of brain hubs be at the base of those cognitive capacities distinguishing humans from other animal species[147], while

an abnormal rich-club organization has been found in patients suffering from several pathologies[148–153]. Also social networks have rich-clubs. Resources, for example, are usually not homogeneously distributed and a subset of prominent individuals controls a large part of all the resources[154–156]; scientific collaborations, in terms of the co-authorship of papers, tend to be stronger between renowned scientists[157]; also face-to-face interactions occurring over time are more frequent and more stable than expected by chance among individuals which have more interactions[158]. The rich-club phenomenon can be found also in man-made networks, such as the Internet[139]. Interestingly, there are also real systems lacking a rich-club organization, such as protein-protein interaction networks, where it is claimed[157, 159] that hub proteins are mainly devoted to very specific functions.

However, as we have seen in the previous chapters, the degree is a local measure of importance of a unit in the system and the exchange of information does not always take place along shortest paths, so that a complementary process-driven approach is mandatory to gain a deeper understanding not only of the structure but also of the function of complex systems. In this chapter we present a functional definition of rich club, called *functional rich-club*, as a complementary statistics to the "usual" *structural rich club*, that we are now going to introduce.

4.1 The structural rich club

The structural rich-club phenomenon has been originally quantified through the spectrum of the rich-club coefficient[139], the density of connections among nodes with degree larger than k:

$$\phi(k) = \frac{2E_{>k}}{N_{>k}(N_{>k} - 1)} \tag{4.1}$$

where $N_{>k}$ and $E_{>k}$ are, respectively, the number of nodes with degree larger than k and the number of edges between them. Later, it has been shown[157] that a proper normalization of $\phi(k)$ is mandatory, since $\phi(k)$ alone does not allow filtering out the effect of degree-degree correlations. Normalized descriptors of the rich-club phenomenon have been introduced in[157, 160], dividing the coefficient in (4.1) by the average rich-club coefficient of the uncorrelated or the maximally random network

with the same degree sequence of the network under study

$$\rho_{\text{unc}}(k) = \frac{\phi(k)}{\phi_{\text{unc}}(k)} \tag{4.2}$$

$$\rho_{\text{ran}}(k) = \frac{\phi(k)}{\phi_{\text{ran}}(k)}. \tag{4.3}$$

There are other possible formulations of the normalized structural rich-club[161] and of its generalization to weighted[154, 162–164], hierarchical[165] and temporal[158] networks, but the key point pooling all these descriptors together is the need to distinguish the case where nodes with a lot of (or strong) connections have more links between them just by chance from the case in which hubs have indeed, an intense connectivity giving them, e.g., the control over resources flowing in the system, or facilitating the rapid exchange of information among them.

And here we ask: does a topological, or structural, rich-club organization translate seamlessly into function? Vice-versa, does a functional rich-club organization always need its structural counterpart? To answer these questions, we tackle the rich-club phenomenon using our diffusion geometry framework, presented in Ch. 1. We propose a functional rich-club which is based on the idea that a rapid information exchange among rich nodes is not only facilitated by the presence of many, direct link between them, but also from the global connectivity of the network. For instance, bridge nodes with low degree placed between hubs, may be overlooked in the structural case, whereas using an averaging process, like the diffusion, we can account also for longer paths. Furthermore, a small set of rich-nodes may form a strong structural rich-club, regardless of the presence or not of other meso-scale structures. Instead, in the functional setting these structures may be competing at some scale, leading to a lack of a functional rich-club organization.

4.2 The functional rich-club

The idea is to use the diffusion distance to determine if the rich nodes in a network form a functional club: if the diffusion distances between rich nodes are, on average, smaller than the average distance between any pair of nodes in the network and also smaller than what is expected by chance, then there is a functional rich-club.

Let us begin with a brief recap—the notation is the same as in chapter 1, as well as all the definitions. Given a network $G = (V, E)$, the diffusion distance[5] between

two nodes $i, j \in V$ is defined by eq. (1.4)

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

where $t > 0$ is the diffusion time and $\mathbf{p}(t|i)$ indicates the probability vector whose $k-$th component $p_k(t|i)$ is the probability of finding the random walker at node k at time t given that (i) it started at node i at time $t = 0$ with probability 1 and (ii) it follows a continuous-time classical random walk on the network G. In discrete time a classical random walk is characterised by the transition matrix $\mathbf{T} = \mathbf{D}^{-1}\mathbf{A}$, while in continuous time the random walk is controlled by the normalized Laplacian as $\tilde{\mathbf{L}} = \mathbf{I} - \mathbf{T}$ and evolves according to the forward equation $\dot{\mathbf{p}}(t) = -\mathbf{p}(t)\tilde{\mathbf{L}}$. With the initial condition $\mathbf{p}(0) = \mathbf{e}_i$, the transition probability at time t is given by $p_k(t|i) = \left(e^{-t\tilde{\mathbf{L}}}\right)_{ij}$. Recall also that $D_t(i,j) \in [0, \sqrt{2}]$ for all $i, j \in V$ and $t \geq 0$, and that if G is connected $D_t(i,j) \to 0$ as $t \to \infty$[7].

Studying the distribution of these mutual diffusion distances, we are able to tell if rich nodes are indeed nearer to each other than expected by chance, accounting also for the rest of the nodes in the network. To remove the effect of the whole network we standardise the diffusion distances as follows: firstly, we compute the average $\mu(D_t)$ and standard deviation $\sigma(D_t)$ of the diffusion distances over the whole network and then compute the pairwise standardised distances

$$\tilde{D}_t(i,j) = \frac{D_t(i,j) - \mu(D_t)}{\sigma(D_t)}. \tag{4.4}$$

Now, $\tilde{D}_t(i,j)$ is zero on average with unitary variance, and may take negative values. As a matter of fact, if $\tilde{D}_t(i,j) < 0$ then i and j are closer to each other w.r.t. the average pairwise distance of the network. Moreover, the magnitude $|\tilde{D}_t(i,j)|$ tells us how many standard deviations far away this distance is from the mean network distance. Henceforth, we drop the "tilda-notation" for the standardized distances, since there is no ambiguity being all distances standardized w.r.t. the network distribution.

For each node degree k we consider the subset of distances between nodes with degree at least k

$$\{D_t(i,j) : i, j \in G, k_i, k_j \geq k\}$$

and compute the average over this set, which we henceforth indicate as $\mu\left(D_t\Big|_{\geq k}\right)$ and refer to as the average core diffusion distance, or simply the core distance of G at level k. Observe that $\mu\left(D_t\Big|_{\geq k}\right)$ is not a statistics, since $\left\{\tilde{D}_t(i,j) : i, j \in G, k_i, k_j \geq k\right\}$ is not a random sample, instead it is a measure of the diffusion closeness of the nested

sub-networks of G. If the average core diffusion distance is negative, then the rich nodes are, on average, closer to each other than any random pair of nodes in the network. To exclude that this closeness is due only to the high degree of the nodes, we need to compare this average distance to the same group of nodes in a null model.

Let $G_1, \ldots, G_n$ be a sample from the configuration model[166] with the same degree sequence of G. For each realization G_i, we compute the diffusion distance at the same time scale and repeat the same standardising procedure as in (4.4). We then collect the set of core distances for each G_i, put them together in a unique large sample, and finally compute its sample mean $m\left(D_t^{\mathrm{cm}}\big|_{\geq k}\right)$ and standard deviation $s\left(D_t^{\mathrm{cm}}\big|_{\geq k}\right)$.

The functional rich-club coefficient $\mathrm{FRCC}_{t,k}$ is finally defined in terms of a standardized distance from the configuration model. Since we would like to observe positive values when the network G under study has smaller core distances than the null model, we take

$$\mathrm{FRCC}_{t,k} = \frac{m\left(D_t^{\mathrm{cm}}\big|_{\geq k}\right) - \mu\left(D_t\big|_{\geq k}\right)}{s\left(D_t^{\mathrm{cm}}\big|_{\geq k}\right)}. \tag{4.5}$$

which quantifies how far the average network core distance is from the average configuration model core distance, in units of the standard deviation of the configuration model core distances.

One advantage of using standardized scores is that they can be compared without worrying about scales: we can compare the $FRCC_{t,k}$ at different diffusion times and, if we standardize the structural rich-club coefficient, we can also compare the two rich-club organizations. The Z-score for the structural rich-club coefficient is evaluated similarly (but with inverted sign)

$$\phi^Z(k) = \frac{\phi(k) - m\left(\phi^{\mathrm{cm}}(k)\right)}{s\left(\phi^{\mathrm{cm}}(k)\right)}. \tag{4.6}$$

A model of functional rich-club

We start our analysis with a possible model of a synthetic (unweighted and undirected) network with defined and controllable a functional rich-cub organization. A network with a structural rich club is easily build starting from any network with hubs, that is with a heterogeneous degree distribution, adding a lot of edges (or all possible edges) among these high-degree nodes. For the functional rich-club, however, we expect that it will not be sufficient to manipulate only the local connectivity of hubs, but that we

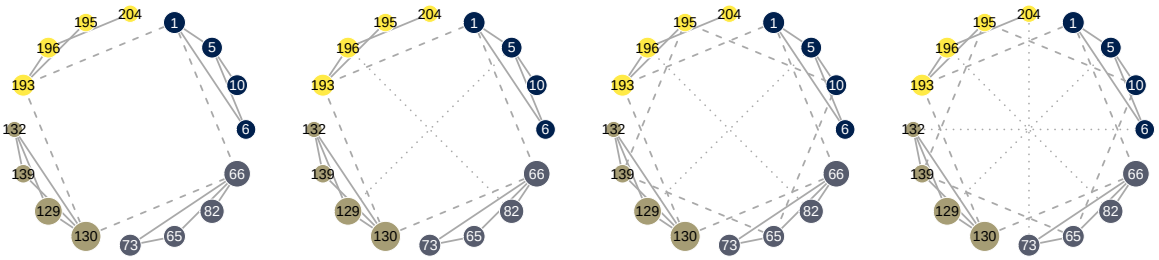

Figure 4.1: Ring-and-crosses connectivity pattern between rich nodes

will need to add more walks of different lengths between rich and less-rich nodes too.

We start with a network with hubs or, to be more precise, with four sub-networks $G_1, \ldots, G_4$ with 64 nodes each, generated from the Barabási–Albert (BA)[102] model (with $m = 2$) that are then loosely connected at step 1. The four sub-networks can be seen as communities of the final network; nodes inside each community are ranked in decreasing degree order, and we call the top four ranking nodes in each group *the rich nodes*. The inter-community edges between rich nodes are placed according to their rank following a "rings and crosses" pattern, as shown in Fig. 4.1. For rank 1 or 3, nodes with the same rank are connected in a *ring*, i.e., the first (third) ranking node in community G_i is connected to the first (third) ranking node in G_{i+1} for $i = 1, 2, 3$ and G_4 is connected to G_1, while for ranks 2 and 4 we use a *cross* connection pattern, the node with second (fourth) highest degree in community G_1 is connected to its homologous in G_3 and the one in G_2 to G_4. Finally, five random edges link the five lowest degree nodes of each pair of distinct communities, more precisely a random edge between communities G_i and G_j is obtained sampling, uniformly at random with replacement, each endpoint from the respective community. Multiple edges, if any, are removed. This concludes the first building step.

At the second and third steps, inter-rank links between rich nodes are added in order to create paths of length larger than one between rich nodes, as shown in Fig. 4.2. At step 2, the first-ranking node of each group G_i is linked at random with (i) two fourth-ranking nodes outside G_i, (ii) one third-ranking node of a group chosen at random in $\{G_1, \ldots, G_4\} \setminus \{G_i\}$, indicated henceforth as $G \setminus G_i$, and, similarly, one second-ranking node in $G \setminus G_i$. In the same manner, at step 3, the second-ranking node of each group connects randomly to two third-ranking nodes and to one fourth-ranking node of the other groups.

Finally, at the step 4 we compute the 75−th and 90−th percentiles of the degree distribution, corresponding to $k_{.75} = 4$ and $k_{.9} = 8$ and we place ten edges between ten nodes with a degree $k > k_{.9}$ chosen at random inside group G_i to ten nodes

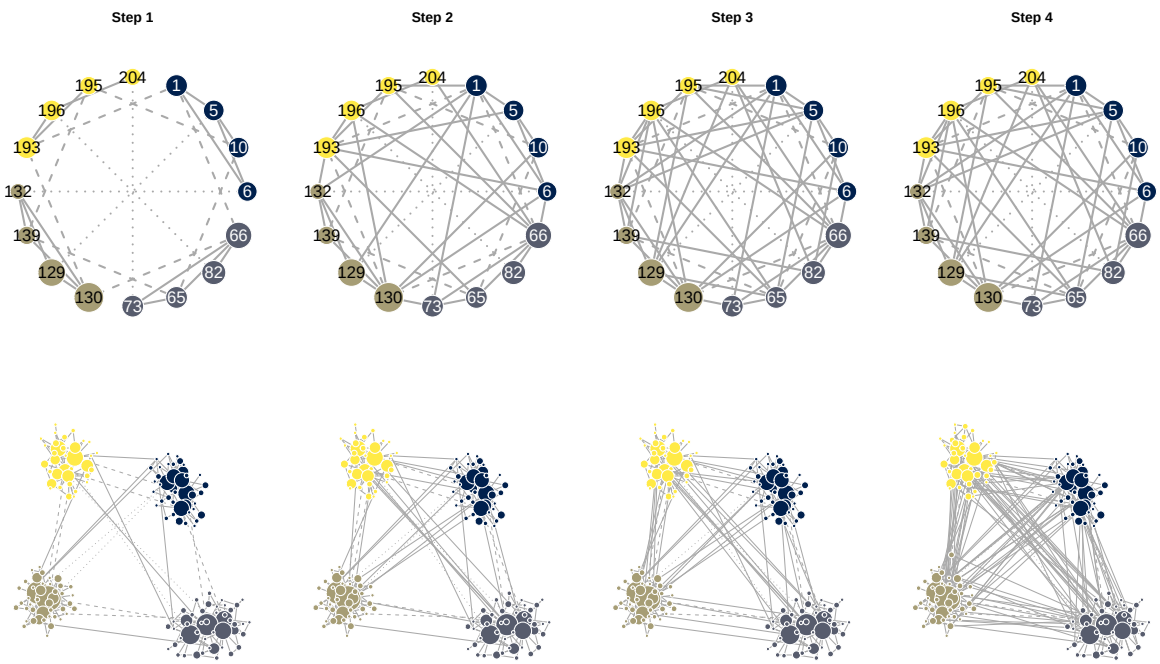

Figure 4.2: Model of a functional rich-club organization: moving from weakly connected sub-networks to more densely connected rich and mildly-rich nodes. Connectivity of the rich nodes and their neighbours in the four building steps of the model. The edge density (fraction of possible links) among the rich nodes increases from 0.23 at step 1, to 0.47 at step 3 (in the last step no new link is added between rich nodes).

with a degree $k_{.75} \leq k \leq k_{.9}$ chosen at random in $G \setminus G_i$. The average degree of G at the different steps varies from $\langle k \rangle \approx 4.2$ to $\langle k \rangle \approx 4.9$ and the edge density among rich nodes, i.e. the ratio between the number of existing edges over all possible $\frac{16 \cdot (16-1)}{2} = 120$ edges, goes from 23% at step 1, to 47% at step 3.

Next, in Fig. 4.3 we show the functional rich-club coefficient $\text{FRCC}_{t,k}$, eq. (4.5), for varying degree k and diffusion time t and, for the structural part, we show in Fig. 4.4 both the Z-score of the original rich-club coefficient $\phi(k)$, indicated as $\phi^Z(k)$ and computed according to (4.6) and the rich-club ratio $\rho(k)$ of (4.2).

As we said, the network has a very strong organization of its nodes in four equally sized ($\frac{N}{4} = 64$) communities. At the beginning these communities are very weakly connected, through random edges between low-degree nodes and through regular connection patterns (rings and crosses) between the four richest nodes in each group, indicated henceforth as the rich nodes. This is enough for observing a (weakly significant) structural rich-club for $k > 15$, according to $\rho(k)$ in Fig. 4.4, while, functionally, the average distance between rich nodes is significantly larger compared to the configuration model, especially for $4 < t < \frac{N}{4}$. Adding more links between the rich-nodes (steps 2-3) strengthens the structural rich-club and brings the rich nodes

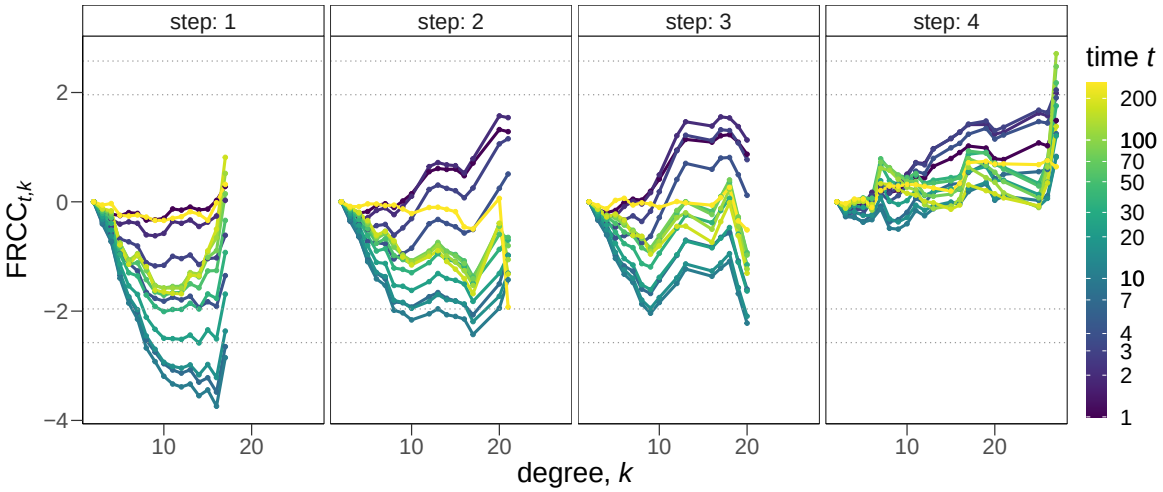

Figure 4.3: The emergence of a functional rich-club. The functional rich-club coefficient of the synthetic model during its construction. Dashed lines enclose 95% and 99% of the standard normal distribution. Points outside these ranges are sufficiently extreme deviations from the mean of the configuration model, which allow as to claim that G has a strong functional rich-club. At step 1 five inter-communities random edges are placed between low-degree nodes for each pair of communities; additionally rich-nodes of the same rank but belonging to different groups are linked through "rings and crosses". Here, the strong meso-scale organization of the network causes diffusion distances between nodes of degree $k \geq 10$ to be significantly larger than expected and this reflects in the absence of a functional rich-club, or even the presence of an anti-rich-club. In steps 2-3 we add inter-rank links between rich-nodes: this increased connectivity among hubs of different groups has a strong effect on the local functional organization of the network and nodes with a degree higher than $k = 10$ start to belong to a weak functional rich-club. When, in step 4, mildly-rich nodes (i.e. nodes with a degree k in the range $[4, 9]$) are connected to nodes with degree larger than $k = 9$, a strong functional rich-club emerges at different scales.

nearer in the diffusion space at small time scales, i.e., $t < 4 \approx \langle k \rangle$, the average network degree. Finally, a functional rich-club for $k \geq 25$ emerges when we connect rich and mid-rich nodes.

It is also interesting to observe that for small time scales, say $t < 5$, the structural and functional Z-scores display a similar but not identical behaviour. Thinking about the embedded chain of our continuous-time random walk, i.e., a classical discrete-time random walk on the network, one observes that for $t = 1$ a random walker can only reach the first neighbours of the initial node and so, one could expect the two rich-clubs to be the same. However, a more attentive look reveals a subtle but non negligible difference: if the structural rich-club coefficient $\phi(k)$ is proportional to $\sum_{k_j > k} k_j$, disregarding of the rest of the network, the average distance between nodes with degree larger than k always accounts for all edges going out from those rich nodes. In particular, if two random walkers starting in two distinct rich nodes have both a

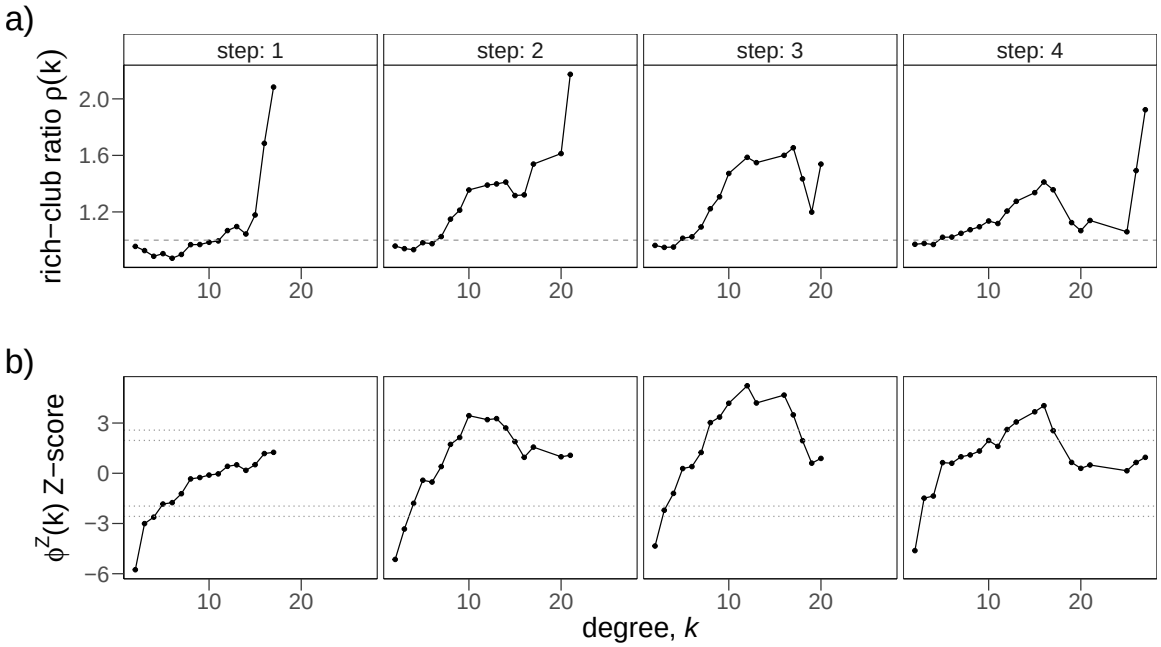

Figure 4.4: Structural rich-club in the model network. **a)** Z-scores of the structural rich-club coefficient, $\phi^Z(k)$ and **b)** the ratio $\rho(k)$. Dashed lines enclose 95% and 99% of the standard normal distribution. Points outside these ranges are sufficiently extreme deviations from the mean of the configuration model, which allow as to claim that G has a strong structural rich-club. There is a clear structural club of mid-rich nodes, i.e., for degree $k \in [8, 12]$, at all steps (only weakly at step 1, according to ϕ^Z). In step 4, adding connections between rich and mild-rich nodes has a negligible effect on the structural rich-club ordering of nodes with degree $k \in (10, 20)$.

non zero probability of going into a third non-rich node, then this will decrease their reciprocal diffusion distance. On the other side, if two rich nodes are directly linked but have otherwise disjoint first neighbourhoods, then the posterior distribution of the random walker being in any node after one step will be almost localized in the two neighbourhoods and, consequently the diffusion distance of the two node will be high.

This simple toy network suggests that the global network structure plays an important role in shaping/conditioning the information flow among hubs: if these belong to distinct communities in which they are, having high degree, deeply-rooted, despite the presence of direct links between them, the diffusion is hindered, resulting in higher distances in the diffusion space w.r.t. the null model, where the meso-structures are destroyed. To answer our previously asked questions, we have shown that a network can have a significant structural rich-club and lack the functional rich-club, Figs. 4.3-4.4 steps 2-3. Vice-versa, in presence of bridges and many short (even if not direct) paths between hubs, the diffusion among them may be eased sufficiently enough to form a functional rich-club., Fig. 4.3 steps 4.

Following our usual analysis roadmap, we now consider more realistic network models and some real-world networks, keeping in mind the mechanisms that allow or hinder the presence of a diffusion-driven rich-club—short paths between rich and mid-rich nodes vs. a strong community structure.

Functional rich-clubs in realistic and real networks

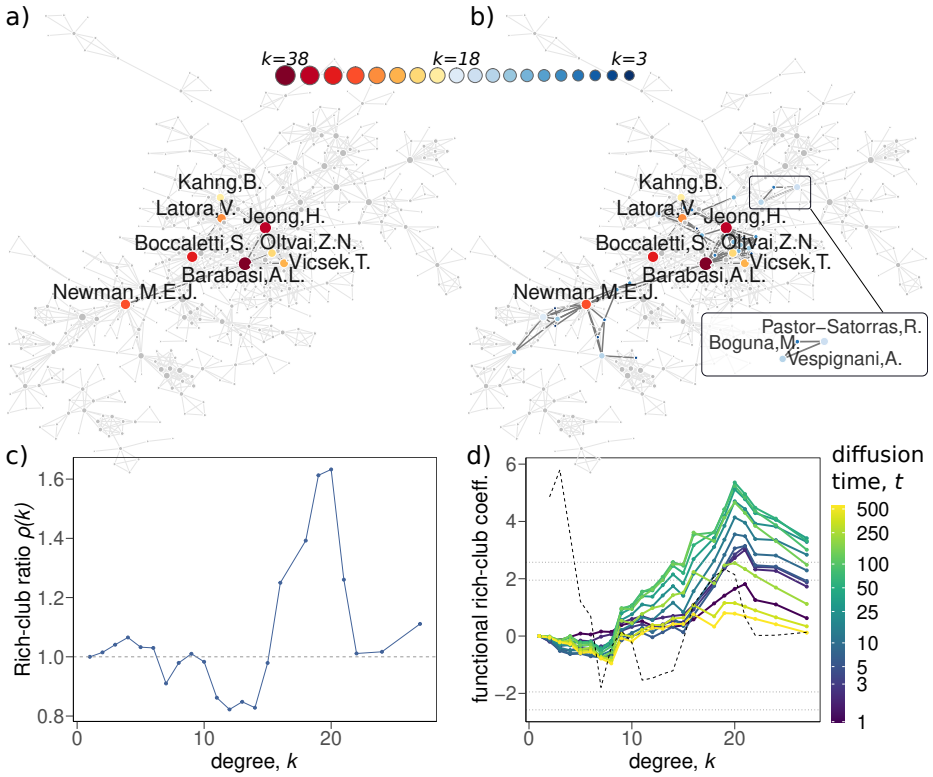

Figure 4.5: Network of 2010 Network scientists. a) Its structural rich-club, where nodes with a degree equal to or larger than 20 are coloured: and labelled. b) All the nodes which are in the same (metric) neighbourhood of the rich-nodes. These nodes are not rich themselves, but contribute to the diffusion among rich nodes. c) The rich club ratio $\rho(k) = \frac{\phi(k)}{\langle \phi(k)_{\mathrm{cm}} \rangle}$. For $k > 15$ the structural rich-club coefficient is larger than 1, indicating the presence of a rich-club, with a peak for $k = 20$, while for larger values of the degree the rich-club is not so evident. d) The functional rich-club coefficient as in (4.5), as a function of the degree, k, and diffusion time, t; the dashed line corresponds to the Z-score of the structural rich club $\phi^Z(k)$. Dashed horizontal lines delimit the ranges $[-1.95, 1.95]$ and $[-2.575, 2.575]$ which enclose resp. 97.5% and 99% of the standard normal distribution. There is a strong evidence of a functional rich-club for $k > 15$ at different time scales t. For $k = 20$ the average core diffusion distances are more than three standard deviations smaller than the distances in the cores of the configuration model. Differently from its structural counterpart, the functional rich-club is clearly present also for richer nodes, i.e., $k > 20$.

From previous studies[154, 157] we know that networks representing social and col-

laboration relationships between individuals show the structural rich-club phenomenon, so we look at the results of the functional rich-club analysis on the unweighted 2010 Networks Scientists Network[167, 168], a co-authorship network with 522 nodes, where two scholars are linked if and only if they wrote at least one paper together, which are summarized in Fig. 4.5. At a local scale, $t = 1$, there is no significant functional organization of the rich nodes into a club, as shown in panel d), although for $k = 20, 21$ the average distance between nodes of degree at least 20/21 is almost two times smaller than expected by chance (i.e., according to the average of the configuration model). Also the structural coefficients $\rho(k), \phi_Z(k)$ detect a strong rich-club at $k = 19, 20$, as it can be seen in panels c) and d)–dashed black line–respectively. As time increases, the diffusion between rich nodes ($k \geq 18$) is significantly supported by the presence of longer paths between them, leading to a strong evidence of the functional rich-club phenomenon. In the top panels of Fig. 4.5 we show a) the structural and b) the extended functional rich-clubs for $k \geq 20$. More specifically, in Fig. 4.5-b) we fix $t = 2$ and highlight all the nodes in the network which are not further than the maximum observed distance in the rich-club, i.e., $\max\left(D_t\big|_{\geq 20}\right)$, from each rich node. We call this the metric neighbourhood of the rich nodes. Additionally to the 8 rich nodes, this neighbourhood contains other 30 nodes with varying degree, which may belong to small and dense communities of or connecting the rich-nodes, as for instance the nodes around Newman or the community to which Barabási, Jeong, Oltavai and Vicsek belong. Interestingly, there is also a small clique, consisting of Boguña, Pastor-Satorras and Vespignani, that is topologically, at one-step, disconnected from the rich nodes, the completely labelled neighbourhood sub-network is shown in Fig. 4.6. To summarize, in the diffusion space the rich nodes are close to each other not only because of their direct connections, but also for the presence of different small structures (bridges, cliques and dense communities) supporting diffusion.

The same holds for other real networks, such as the connectome in the Drosophila's medulla[169, 170] and the world-wide airports network[8], while the protein-protein interaction (PPI) network of the Plasmodium falciparum[171, 172] has generally no rich-club, as shown in Fig. 4.7-a)–c).

These networks, whose details and references are reported in Tab. 4.1, have been analysed along network models. Figure 4.7-d)–f) shows the functional rich-club coefficients for the Barabási–Albert (BA)[102], the Erdős–Renyi (ER)[173] and the Lancichinetti–Fortunato–Radicchi (LFR)[174] model; others can be found in the Appx.A.5. All the synthetic networks have $N = 256$ nodes and their parameters have been chosen to result in an average degree $\langle k \rangle \approx 12$ (see the Appx.A.5 for details);

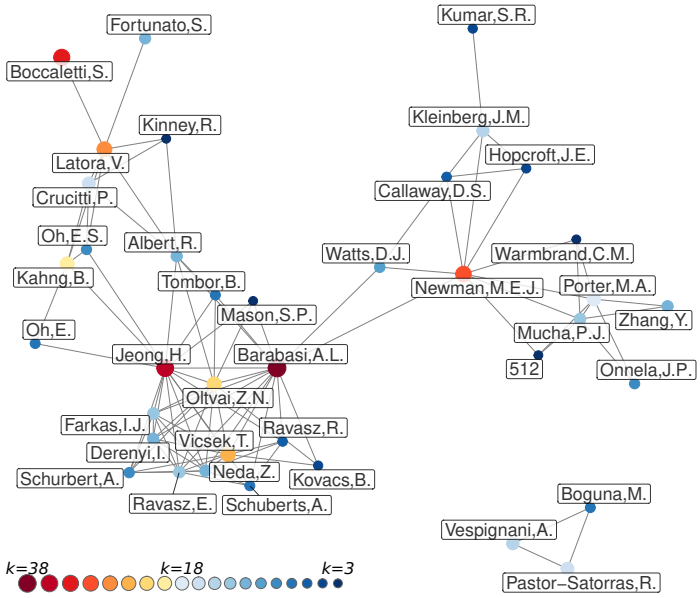

Figure 4.6: Metric neighbourhood of the rich-club of the Network Scientists 2010 network. Nodes size and colour depends on their degree: nodes with a degree $k \geq 20$ (rich nodes) are coloured in shades from yellow to red, while nodes coloured in shades of blue are the nodes belonging to the metric neighbourhood of rich nodes, which is the intersection of the balls of radius $\max\left(D_t|_{\geq k}\right)$ centred in each rich node. Blue nodes are not rich, nonetheless, they contribute to the proximity of rich nodes in the diffusion space, as bridges or by forming dense communities across several rich nodes.

notwithstanding, they have very distinctive characteristics, which are reflected in their functional organization. BA networks have a heterogeneous degree distribution with hubs and these form a functional rich-club. As expected, the nodes in ER networks are neither closer to each other nor mutually more distant than by chance. Finally, the LFR model, which is characterized by the presence of both hubs (heterogeneous degree distribution) and communities with different sizes, further confirms our conjecture that networks with a clear community structure lack a functional rich-club organization. On the contrary, we can see in Fig. 4.7-f) that as nodes become richer, their average core distance grows more than expected by chance.

Table 4.1: Real networks. Size and average degree of some real networks. We considered different types of biological networks: a promote—the protein-protein interaction (PPI) network of the malaria parasite Plasmodium falciparum—and a connectome. An instance of infrastructural network is given by the world-wide airports network. This is the reduced version, with 965 airports out of the over four thousand airports of the complete network, proposed in[8].

	Netsci 2010	P. falciparum PPI	Dros. connectome	Airports
n. of nodes	552	1179	1170	965
n. of edges	1318	2481	8905	33349
avg. degree $\langle k \rangle$	4.8	4.2	10.1	69.1

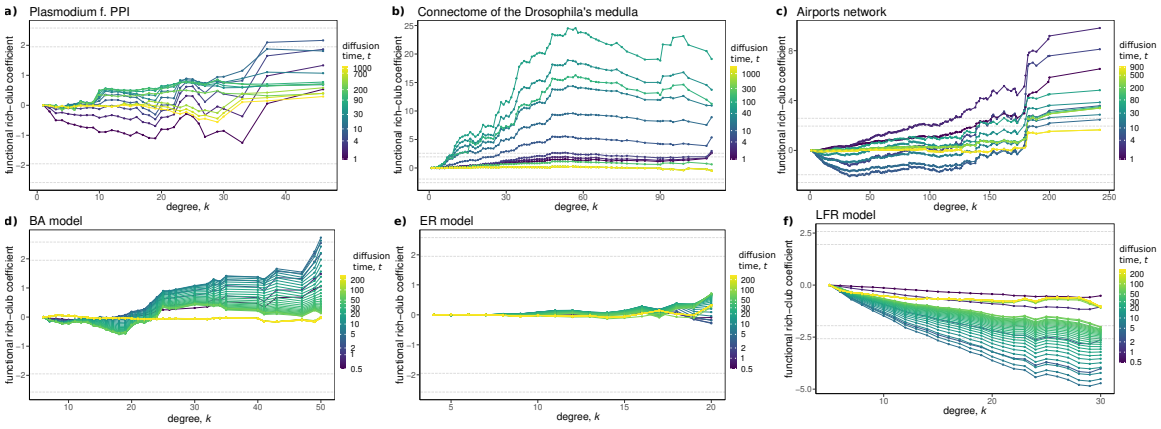

Figure 4.7: The functional rich-clubs of real and synthetic networks. See Tab. 4.1 for details on the datasets. a) The protein-protein interaction network (PPI) of the P. falciparum has no persistent functional rich-club organization, while b) the connectome in the Drosophila's optical medulla and c) the worldwide air transportation network have a clear functional rich-club at different scales, i.e., diffusion times t. Synthetic networks are generated from three models: d) Barabási–Albert (BA), e) Erdös–Renyi (ER) and f) Lancichinetti–Fortunato–Radicchi (LFR). These networks have 256 nodes each and an average degree around 5, nevertheless their different structural features are reflected in their functional rich-club organizations: The BA network has a significant functional rich-club at the highest values of k, while the ER network is not different from the null model. The LFR has a clear community structure and lacks the functional rich-club organization; on the contrary, as nodes become richer their average core distance grows more than expected by chance.

To conclude, we have shown that not only the absence of a structural rich-club, but also the presence of clear meso-scale structures, does not facilitate or even hinders the diffusion between rich nodes. Indeed, functional organizations of nodes into communities, cores and rich-clubs are competing at some scale. Furthermore, the frequent claim that systems which need to integrate information processed in segregated subsystems display the rich-club phenomenon, finds here further support.

4.3 Discussion and conclusion

Many empirical networks are not static and connections are purposefully created or destroyed to achieve a particular function, e.g. cognitive capacities emerge from neural signalling in our brain, local perturbations in the international trade network affect far away countries, interactions between people shape not only our daily life but have consequences on the spreading of viruses, etc. The interplay between structure and function has to be taken into account when ascribing a particular functional output to an observed structural pattern. Although the rich-club phenomenon is found in

different types of systems and is considered responsible for the rapid exchange and integration of information, its definition is purely structural and does not allow us to specify the actual communication mechanism. In this work, we made a first attempt to fill this gap by tackling the rich-club phenomenon from the functional perspective of diffusion geometry, asking whether rich nodes are closer in the diffusion space than expected by chance and how this relates to the structural rich-club organisation. From the analysis of our toy model and synthetic networks we have shown that the structural rich-club organization does not translate, seamlessly, into function, because the existence of communities can hinder the diffusion between rich nodes despite the one-link paths connecting them. Remarkably, we have shown that a strong functional rich-club may emerge without (e.g., the airports network for large degrees), or with a weak (2010 Network Scientists) structural counterpart, because the diffusion geometry at not too local time scales integrates the structural information of longer paths so that these contribute to an efficient diffusion between nodes. With these results, we are foreseeing the potential of this framework if extended to other communication mechanisms, which lie in between routing and broadcasting and are important to many real systems.

Chapter 5

Conclusion

Networks, despite their minimalistic aspect as graphs $G = (V, E)$, are complex objects in which pairwise interactions (the edges E) between nodes (elements in V) may produce unexpected outcomes, emergent behaviours at the global network scale, such as the small-world and rich-club phenomena, effective transport, efficient communication, synchronisation and consensus, cognition and life. Networks arise from the most disparate areas of science and of the real-world, we are part of networks and networks condition our lives. For this reason, tackling the analysis of networks from different perspectives is of the utmost importance and enriches our understanding of our complex world.

The focus of this dissertation is the network diffusion geometry. This is an effective geometry induced by random walks and diffusion processes on networks and hence does not require any hypothesis on the metric space or manifold the nodes might live in. In this framework different random walk and diffusion dynamics correspond to different exploration or communication strategies and the resulting geometries reflect not only structural characteristics of the network but also the peculiar dynamical features of the process. From the diffusion geometry perspective, the dynamics changes the graph: every time-homogeneous Markov chain (MC) on finite states can be represented as a graph and different transition rules on a given network result in different MC graph representations, which may seem a weakness of the diffusion geometry framework. On the contrary, this is one of its major strengths. Think about a road network, which effectively transports people and goods in a city when the flows are in a certain range, but as soon as traffic increases it is not able any longer to carry on its function as efficiently as before. The flow changes the output of the structure-function interplay. Network analysis cannot be done isolating structure and function, because the two complement each other and are both necessary for understanding the final functions,

meant as the system specific tasks, of the system.

Random walks and diffusion are only a class of processes that can take place on networks, nevertheless they are good approximations of many dynamics, among which synchronisation and consensus, with their different flavours they represent different information exchange strategies and least but not last, are mathematically appealing since they intersect with spectral theory and allow us to rephrase structural/dynamical features in probabilistic terms. For instance, node centrality can be seen as ease of diffusing but also as a high probability that a random walker starting in a central node runs into other random walkers walking on the network[15]; functional damage increases the dispersion of the nodes in the diffusion space[44]; ultimately, the distance between nodes is defines as a distance between probability vectors[5, 7].

In this thesis I tried to highlight that when we represent a system as a network, a hypergraph, a multilayer network, a simplicial complex, we make two basic assumptions: (i) that we are in the discrete and (ii) that the sparse pairwise or higher-order interactions shape the observed outcome of the system (its behaviour, its function). The interactions are not static relations, they are the mean for something to happen, i.e. for an exchange of information, of a signal, of some physical quantity. In brief, for communication. Hence, understanding how different basic communication strategies shape networks (their functional communities[5] and rich-clubs, process-motifs[12], global efficiency[16] and node centrality[15]) may move us closer to the comprehension of the systems' (dis-)function.

To this end, future perspective of this work and of the whole process-driven geometry framework are two-fold. From one hand, we are far from a full understanding of the basic communication dynamics in many empirical systems—e.g., in the brain signal do not follow shortest-paths but they do not diffuse either—and hence the structure, function, and geometries of these systems should be studied from different perspectives. The framework developed in this thesis provides a set of tools, which can lead to new insights in practical applications and, vice versa, could be largely improved by the field-specific knowledge coming from data and real systems. On the other hand, the space for mathematical improvement of this framework is unlimited: a thorough comparison of the diffusion geometry framework with the geometric structures induced by other random walk statistics (commute and first hitting times) as well as other (diffusion) kernels or electrical networks analogies, is still missing. From a purely mathematical perspective, and as a very personal curiosity, it would be interesting to investigate, from the point-of-view of information geometry, the diffusion space of networks: do random walks map different topologies to identifiable (possibly in

a non-parametric fashion) probability distributions? Furthermore, as each network-driven process may induce an effective geometry, scientists might profit from a general framework for defining, analysing, and comparing them.

Appendix A

Supplementary Informations and Materials

A.1 Continuous-time Markov chains

In this appendix, we will provide the basic notions and results concerning continuous-time Markov chains. The main reference is[40].

A continuous-time Markov chain (CTMC) can be constructed in different ways[40]. Here we will describe the Poisson process, a simple CTMC, and provide the definition of general CTMCs in terms of jump chain/holding times, and transition rates.

Let $(\Omega, \mathcal{E}, \mathbb{P})$ be a probability space. A right-continuous process $(X_t)_{t>0}$ with values in a countable set I is a family of random variables $X_t : \Omega \to I$ such that the sample–paths $t \to X_t(\omega)$ are right-continuous $\forall \omega \in \Omega$. The probability of events of a right-continuous process are then completely determined by its finite-dimensional distributions $\mathbb{P}(X_{t_0} = i_0, \ldots, X_{t_n} = i_n)$, with $0 < t_0 < \cdots < t_n$ and $i_0, \ldots, i_n \in I$ (Kolmogorov theorem). Let $T_0, \ldots, T_n$ be random variables over $(\Omega, \mathcal{E}, \mathbb{P})$ such that

$$T_0 = 0$$
$$T_{n+1} = \inf\{t \geq T_n : X_t \neq X_{J_n}\}$$

for $n = 0, 1, \ldots$. Define also

$$S_1 = T_1, \ldots, S_n = T_n - T_{n-1}.$$

Variables T_i are the random times at which events occur, while S_i are the waiting (or holding) times (see[40] for details on explosion times, which are not considered here).

The discrete-time process $Y_n = X_{T_n}$ is called the *jump process* of $(X_t)_{t>0}$, or the *jump chain* if it is a discrete-time Markov chain, or also the *embedded* chain of $(X_t)_{t>0}$.

Definition A.1 (Poisson process)**.** A right-continuous process $(X_t)_{t>0}$ with values in $\{0, 1, 2, \dots\}$ is a Poisson process of rate λ $(0 < \lambda < \infty)$ if its holding times S_i are independent exponential random variables of parameter λ and its jump chain is given by $Y_n = n$.

A Poisson process can be equivalently defined ([Thm. 2.4.3, 40]) in terms of (i) its jump chain and holding times (as in Def. A.1), (ii) of its increments $X_{t+s} - X_s$ (which are independent), (iii) of its transition probabilities (for each t $X_t \sim \mathcal{P}(\lambda t)$), or in terms of (iv) its transition rate $Q-$matrix:

$$\mathbf{Q} = \begin{pmatrix} -\lambda & \lambda & & \\ & -\lambda & \lambda & \\ & & \ddots & \ddots \end{pmatrix}$$

Recall the definition of a $Q-$matrix[40]

Definition A.2 ($Q-$matrix)**.** Let I be a countable set. A $Q-$matrix on I is a matrix $\mathbf{Q} = \{q_{ij} : i, j \in I\}$ such that:
(i) $0 \leq -q_{ii} < \infty$ for all i
(ii) $q_{ij} \geq 0$ for all $i \neq j$
(iii) $\sum_{j \in I} q_{ij} = 0 \quad$ for all i.
q_{ij} is the rate of going from i to j and $q_i := -q_{ii}$ is the rate of leaving i.

Finally, the Thm. 2.8.2 of[[40] provides as three equivalent ways to define a continuous-time Markov chain (CTMC) on I with generator ($Q-$matrix) $\mathbf{Q}$ and initial distribution μ is a right-continuous process $(X_t)_{t>0}$ such that:
(i) **jump-and-hold definition** for $i, j \in I$ let $J_{ij} = \frac{q_{ij}}{q_i}$ for $i \neq j$ and $J_{ii} = 0$ be the jump matrix corresponding to $\mathbf{Q}$, conditional on $X_0 = i$ the jump chain $(Y_n)_{n \geq 0}$ of $(X_t)_{t>0}$ is a discrete-time Markov chain with transition matrix $\mathbf{J}$ and initial distribution δ_i. Furthermore, for each $n \geq 1$ conditional on $Y_0, \dots, Y_{n-1}$ the holding times $S_1, \dots, S_n$ are independent exponential random variables of rate $q(Y_0), \dots, q(Y_{n-1})$ respectively.
(ii) **(infinitesimal description)** for all $t, h \geq 0$ given that $X_t = i$, X_{t+h} is independent of $(X_s : s \leq t)$ and as $h \to 0$, uniformly in t, the chance that $X_{t+h} = j$ is $\delta_{ij} + q_{ij}h$
(iii) **(transition probabilities/forward equation definition)** for all $n = 0, 1, \dots$

all times $0 \leq t_0 \leq t_1 \leq \cdots \leq t_{n+1}$ and all states $i_0, \ldots, i_{n+1}$

$$\mathbb{P}\left(X_{t_{n+1}} = i_{n+1} \mid X_{t_0} = i_0, \ldots, X_{t_n} = i_n\right) = p_{i_n i_{n+1}}\left(t_{n+1} - t_n\right)$$

where $(p_{ij}(t) : i, j \in I, t \geq 0)$ is the solution of the forward equation

$$P'(t) = P(t)Q, \quad P(0) = I$$

i.e. $p_{ij}(t) = \left(e^{\mathbf{Q}t}\right)_{ij}$.

A.2 Basic results and proofs about graphs

Eigenvalues of the Laplacian matrices

Let $\mathbf{L} = \mathbf{D} - \mathbf{A}$ be the combinatorial Laplacian and let λ be an eigenvalue of $\mathbf{L}$ with corresponding eigenvector $\mathbf{x}$ (w.l.o.g. we choose $\mathbf{x}$ of unit length). Then

$$\begin{aligned}
\lambda = \mathbf{x}\mathbf{L}\mathbf{x}^T &= \sum_{i=1}^{N} x_i \left(\mathbf{L}\mathbf{x}^T\right)_i \\
&= \sum_i x_i \sum_j (\delta_{ij} x_i - A_{ij} x_j) \\
&= \sum_i x_i (k_i x_i - \sum_j A_{ij} x_j) \\
&= \frac{1}{2}\left(2\sum_i k_i x_i^2 - 2\sum_j A_{ij} x_j x_i\right) \\
&= \frac{1}{2}\sum_{i,j} A_{ij}(x_i - x_j)^2 \\
&= \sum_{i\sim j} (x_i - x_j)^2
\end{aligned}$$

where $\sum_{i\sim j}$ indicates the sum over all unordered pairs $\{i, j\}$ of adjacent nodes. This sequence of equalities proves the equality between the scalar product and the *Dirichlet sum of G* in the first chapter of[20]. This equality is used to show that the combinatorial Laplacian is positive semi-definite and hence all its eigenvalues are non-negative.

Average squared diffusion distance (ASDD)

This is part of the Appx. A of[44]. Eq. (1.1) maps nodes in a network to a cloud of points in the diffusion space, whose dispersion can be quantified using a generalized

scalar measure of variance as the trace of their (sample) covariance matrix. Here we show that this is equivalent to computing the ASDD.

Let us fix $t > 0$, so we can drop it from the notation, and call $p_{ij} = \left(e^{-t\mathbf{L}}\right)_{ij}$. Eq. (1.6) becomes

$$\mathcal{D}_t^2 = \frac{1}{2N^2} \sum_{i,j=1}^{N} \sum_{k=1}^{N} \left(p_{ik} - p_{jk}\right)^2$$

while the trace of the sample covariance matrix C corresponding to the N vectors $\left(e^{-t\mathbf{L}}\right)_{i\cdot}$ is

$$\mathrm{Tr}(C) = \frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{N} \left(p_{ik} - \bar{p}_i\right)^2 .$$

The equivalence can be proved using a known property of the variance—also known as variance deformation formula[50]—, but it is also easily proved "by hand" in case particular case of undirected networks, where the Laplacian matrix $\mathbf{L}$ is symmetric. Firstly, observe that for all i $\bar{p}_i = \frac{1}{N}\sum_{k=1}^{N} p_{ik} = \frac{1}{N} =: \bar{p}$ Then,

$$\begin{aligned}
\mathrm{Tr}(C) &= \frac{1}{N} \sum_{i=1}^{N} \left(\sum_{k=1}^{N} p_{ik}^2 - 2\bar{p} \sum_{k=1}^{N} p_{ik} + \sum_{k=1}^{N} \bar{p}^2 \right)^2 \\
&= \frac{1}{N} \left(\sum_{i=1}^{N} m_i^2 - 1 \right)
\end{aligned}$$

where we called $m_i^2 = \sum_{k=1}^{N} p_{ik}^2$ the raw second moment of the $i-$th random vector. Similarly,

$$\begin{aligned}
\mathcal{D}_t^2 &= \frac{1}{2N^2} \sum_{i,j=1}^{N} \sum_{k=1}^{N} p_{ik}^2 - 2 \sum_{k=1}^{N} p_{ik} p_{jk} + \sum_{k=1}^{N} p_{jk}^2 \\
&= \frac{1}{2N^2} \sum_{i,j=1}^{N} \left(m_i^2 - 2(P^2)_{ij} + m_j^2 \right) \\
&= \frac{1}{2N^2} \left(2N \sum_{i=1}^{N} m_i^2 - 2 \sum_{i=1}^{N} \sum_{j=1}^{N} (P^2)_{ij} \right) \\
&= \frac{1}{N} \left(\sum_{i=1}^{N} m_i^2 - 1 \right)
\end{aligned}$$

where we used the undirected network assumption and the semi-group property of $e^{-\tau \hat{H}}$ to write $p_{jk} = p_{kj}$ yielding $\sum_{k=1}^{N} p_{ik} p_{jk} = (P^2)_{ij}$ and then $\sum_j (P^2)_{ij} = 1$.

Finally, two minor observations: firstly, here we use the biased sample covariance, but its unbiased version, with factor $\frac{1}{N-1}$ instead of $\frac{1}{N}$, can also be used provided that the sum of squared distances is also divided by $\frac{1}{N(N-1)}$ instead of $\frac{1}{N^2}$, i.e., the zeros on the diagonal are not counted in the sum. Secondly, the factor $\frac{1}{2}$ in the ASDD definition can also be seen as a re-scaling into $[0, 1]$ of the diffusion distances, which are indeed bounded in $[0, \sqrt{2}]$.

A.3 Multidimensional scaling

Multidimensional scaling (MDS) is a collection of methods aiming to find structure in high-dimensional data, given a $n \times n$ dissimilarity matrix Δ between the observations and the dimension p of the Euclidean space in which data are embedded[131]. The final goal is to find a space configuration of the data points in $\mathbb{R}^p \ni \mathbf{x}_1, ..., \mathbf{x}_n$, such that $\|\mathbf{x}_i - \mathbf{x}_j\| \approx \delta_{ij}$, where δ_{ij} are the original distances and $\|\cdot\|$ is the Euclidean distance in $\mathbb{R}^p$.

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

$$\sigma(X) = \sum_i \sum_j (\delta_{ij} - d_{ij}(X))^2 \tag{A.1}$$

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

In chapters 1 and 3 we use the second approach (MDS solved through majorization) applied to distance or supra-distance matrices and with embedding dimension $p = 3$ (or higher, when explicitly stated). The R package and function used is `smacof::mds`[128].

A.4 Network communication efficiency

On the edge weights of synthetic networks

This is the *Supplementary Note 1* from the Supplementary Materials of[16].

A full network with $N = 30$ nodes has $m = 435$ edges, so we took 30 samples of dimension m from Poisson and power-law distributions with varying parameters. The plots in Fig.2~a) of the main manuscript are obtained combining the probability mass/density functions for some values of the parameters ($\lambda = 1, 12$ and $\alpha = 1.5, 7$) with the boxplots of the samples from those distributions (which constitute the weights of the edges of our synthetic networks). More specifically, for the Poisson distributions we chose a common support $\{1, \dots, 26\}$, where 1[1] is the minimum value over all the samples from the Poisson distribution with parameter $\lambda = 1$ and 26 is the maximum over all the samples from the Poisson distribution with parameter $\lambda = 12$, and then for each $k \in \{0, 1, \dots, 26\}$ we computed the value of the probability mass function $p(k; \lambda)\frac{k^\lambda e^{-\lambda}}{k!}$ for the two parameters values. Since it is difficult to see that the tail of the distribution for $\lambda = 12$ is less-heavy, i.e., the weights distribution is less heterogeneous, we resort to the excess kurtosis of the distribution. This is defined as the kurtosis minus three. The kurtosis, in turn, is the standardized moment of order $n = 4$ of a random variable X, i.e., $\frac{\mathbb{E}\left[(X-\mu)^4\right]}{\sigma^2}$ with μ being the expected value and σ the standard deviation of X. For the Poisson distribution, the excess kurtosis is known to be λ^{-1}. This tells us that the Poisson distribution is lepto-kurtic, i.e., it has fatter tails than any normal distributions (for which the kurtosis equals three). Hence, as λ grows its kurtosis becomes smaller.

The power-law distribution is known to have fat tails, and it is also more evident from the density function plots, that as the power α increases, the tail becomes lighter. Nevertheless, we can verify it mathematically, looking again at the moments of the power-law distribution. Following the definition of the (continuous) power-law distribution provided in[98], i.e., a random variable $W \sim \text{power-law}(\alpha, w_{\min})$ if

$$p(w; \alpha, w_{\min}) = \frac{\alpha - 1}{w_{\min}} \left(\frac{w}{w_{\min}}\right)^{-\alpha}.$$

It is then straightforward to write the moment function for W

$$\begin{aligned}\mathbb{E}\left[W^n\right] &= \int_{w_{\min}}^{+\infty} w^n p(w; \alpha, w_{\min}) dw \\ &= (\alpha - 1) w_{\min}^{\alpha-1} \int_{w_{\min}}^{+\infty} w^{-(\alpha-n)} dw\end{aligned}$$

from which we can see that the moments of order $n \geq \alpha - 1$ are not finite. For

[1]Remember that we shifted the support of the Poisson distribution by 1 to keep the full connectedness of the synthetic networks, by adding 1 to all the values in the samples.

$n < \alpha - 1$ the previous integral becomes

$$\mathbb{E}\left[W^n\right] = \frac{\alpha - 1}{\alpha - 1 - n} w_{\min}^n.$$

The calculations of the moments are longer for the power-law distribution, but are still feasible:

$$\begin{aligned}
\mu =& \mathbb{E}\left[W\right] = \frac{\alpha - 1}{\alpha - 2} w_{\min} \\
\sigma^2 =& \mathbb{E}\left[(W - \mu)^2\right] = \frac{\alpha - 1}{\alpha - 3} w_{\min}^2 - \mu^2 \\
=& w_{\min}^2 \left(\frac{\alpha - 1}{\alpha - 3} - \left(\frac{\alpha - 1}{\alpha - 2} \right)^2 \right).
\end{aligned}$$

The central moment of order four is given by

$$\begin{aligned}
\mathbb{E}\left[(W - \mu)^4\right] =& \frac{\alpha - 1}{\alpha - 5} w_{\min}^4 - 4\mu \frac{\alpha - 1}{\alpha - 4} w_{\min}^3 + \\
&+ 6\mu^2 \frac{\alpha - 1}{\alpha - 3} w_{\min}^2 - 3\mu^4 \\
=& w_{\min}^4 \left[\frac{\alpha - 1}{\alpha - 5} - 4 \frac{(\alpha - 1)^2}{(\alpha - 2)(\alpha - 4)} + \right. \\
&\left. + 6 \frac{(\alpha - 1)^3}{(\alpha - 2)^2 (\alpha - 3)} - 3 \left(\frac{\alpha - 1}{\alpha - 2} \right)^4 \right].
\end{aligned}$$

Finally, we can look its standardized version, which gives us the kurtosis of W: $\frac{\mathbb{E}\left[(W-\mu)^4\right]}{\sigma^4}$. We plotted it as a function of α in Fig.A.1, from which we can see that it decreases as α grows. We can also evaluate its limit for $\alpha \to \infty$, indeed

$$\lim_{\alpha \to \infty} \frac{\frac{\alpha-1}{\alpha-5} - 4\frac{(\alpha-1)^2}{(\alpha-2)(\alpha-4)} + 6\frac{(\alpha-1)^3}{(\alpha-2)^2(\alpha-3)} - 3\left(\frac{\alpha-1}{\alpha-2}\right)^4}{\left(\frac{\alpha-1}{\alpha-3} - \left(\frac{\alpha-1}{\alpha-2}\right)^2\right)^2} = 9.$$

So the excess kurtosis the power-law distribution tends to $6 > 0$, and hence this distribution has always a fatter tail than the standard normal distribution.

Additional figures

In this appendix we collect additional figures comparing the efficiency descriptors introduced in Ch. 2, that we briefly recall here:

- GCE: the global communication efficiency in[16] where physical distances derive

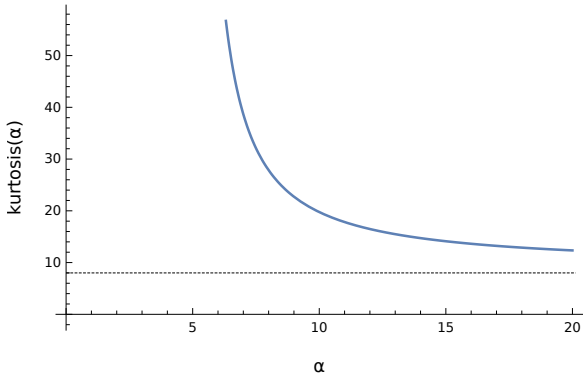

Figure A.1: kurtosis(α) $= \frac{\mathbb{E}\left[(W-\mu)^4\right]}{\sigma^2}$ of the power-law distribution.

from total flows along least-resistance paths.

- E^{T}: the topological efficiency defined in[81].
- E^{LM}: the weighted efficiency descriptor proposed in[81, 83].
- E^{MN}: the max-normalised weighted efficiency $E^{\mathrm{MN}}(G) = \frac{E(G)}{w_{max}}$.
- Total weight (total flow) of the largest connected component (LCC): the sum of the weights in the LCC, normalized in $[0, 1]$.
- Size of the 2$^{\mathrm{nd}}$ largest connected component: the number of nodes in the 2$^{\mathrm{nd}}$ largest connected component, divided by $N = |V|$.
- Number of clusters: the number of connected components in the network, divided by $N = |V|$.

In the boxplots five statistics are shown: the first (Q1) and third (Q3) quartiles, or quantiles of order 25% and 75% (resp. lower and upper box hinges), the median (middle line in the box), the whiskers extend from the smallest to the largest observation in the range $[\mathrm{Q1} - 1.5 \cdot (\mathrm{Q3} - \mathrm{Q1}), \mathrm{Q3} + 1.5 \cdot (\mathrm{Q3} - \mathrm{Q1})]$ Q1 $- 1.5 \cdot$ IQR. Observations falling outside this range are outliers, shown as dots.

Figure A.2 extends Fig. 2.3-a) showing, additionally to the GCE, also E^{MN} and E^{LM}, without limiting to $[0, 1]$ the values on the y-axis. Similarly, Fig. A.3 is the full version of 2.4 without limiting to $[0, 1]$ the values on the y-axis and A.4 is the full version of the inset in the same figure.

We also implemented the variants of the GCE corresponding to the minimum and maximum flow along shortest paths (see eq.(2.8) and (2.10)) and show the results on our synthetic networks ensembles in Fig.A.5.

The two variants, called here GCE$^{\mathrm{min}}$, GCE$^{\mathrm{max}}$, converge faster to 1, since both minimizations (2.8) and (2.10) are less strict than GCE with total flows. Taking the minimum, in particular, may result in values of efficiency spanning a narrow range,

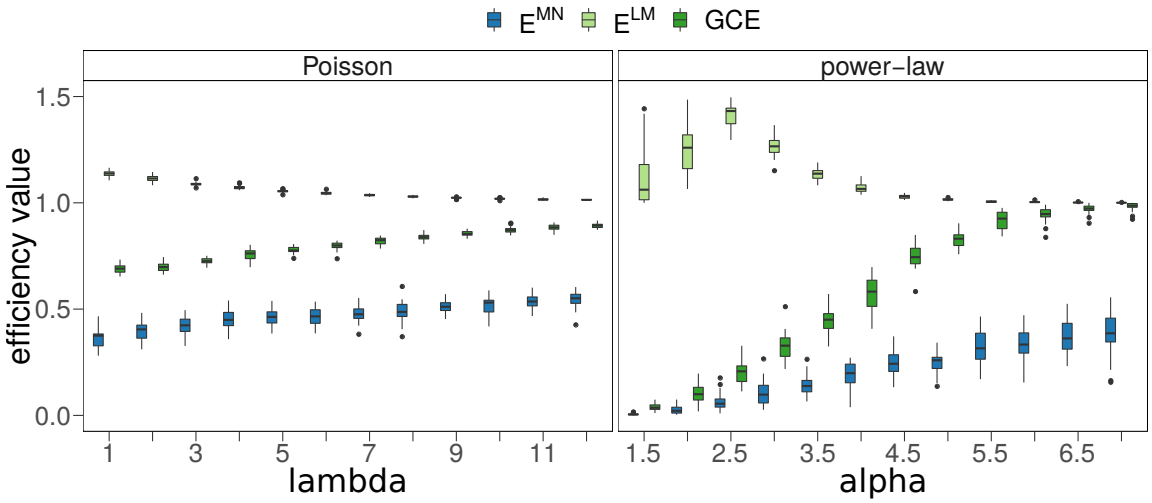

Figure A.2: Comparison of weighted efficiency descriptors for full networks with homogeneous and heterogeneous flows. The descriptors are E^{MN}, E^{LM} and GCE. For each parameter value the descriptors are evaluated for 30 synthetic networks and their distributions are summarized through boxplots. The GCE converges to 1 as the heterogeneity of the weights distributions decreases (i.e., as λ and α increase); E^{MN} remains approximately below 0.5; E^{LM} is not in the range $[0, 1]$. Figure from the Supplementary Materials of[16].

close to 1, with a consequent difficulty in distinguishing networks on the basis of efficiency. Furthermore, in the bottom panel of Fig. A.5 we can see a non-monotone behaviour of $\mathrm{GCE}^{\min}$ which is high for small values of α, when the variance is large (or infinite for $\alpha < 3$, as discussed above). $\mathrm{GCE}^{\max}$ displays, in general, a larger variability than the GCE. Our choice of summing the weights along paths, instead, (i) has a physical meaning in terms of total flow of a subgraph (a path $SP(i, j)$), (ii) it allows us to average the artificial flows matrix with the original flows given by $\mathbf{W}$ and, last but not least, (iii) it is easily worked with in mathematical terms (it simplifies rigorous proofs).

Finally, $\phi_{ij} = 0$ if and only if i, j lie in disconnected components and the ideal network will be disconnected as the original one. In this case, both $d_{ij} = \frac{1}{\phi_{ij}} = \infty$ and the missing links among disconnected components will not produce an under-estimation of the efficiencies of the subgraphs. Of course, if the network is very fragmented the GCE, a global descriptor, will not be very informative. Below, we propose a variant of the GCE, which is most appropriate in this case and in percolation simulations in general.

Our normalisation procedure can also be used to build a slightly modified version of the GCE that plays the role of a weighted integrity descriptor for percolation analysis. Let G_0^{ideal} be the idealized network corresponding to G_0 build as described in our study.

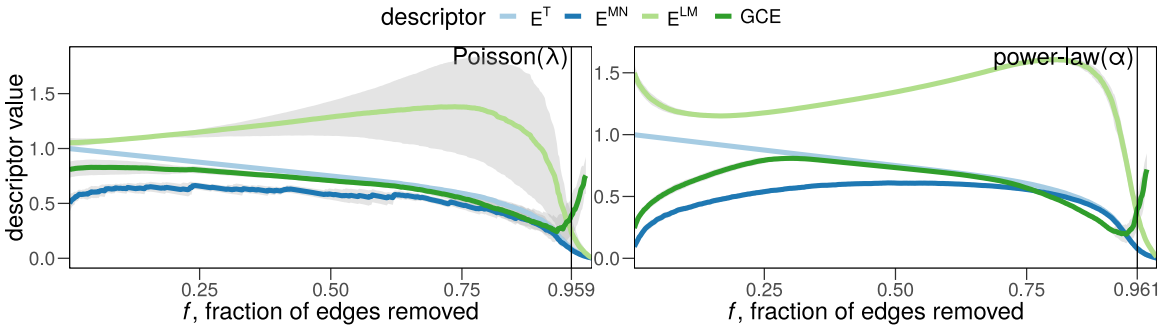

Figure A.3: Targeted bond percolation of an ensemble of synthetic networks with homogeneous and heterogeneous flows. Homogeneous flows are modelled sampling the edge weights from the Poisson distribution, while heterogeneous flows from the power-law distribution. Lines represent the average descriptor value over 30 networks for each model (with $\lambda = 2$ and $\alpha = 2.5$, in particular), while shaded areas the standard deviations. The vertical line indicates the critical f_c corresponding to the maximum of the size of the second largest component. This figure corresponds to 2.4. Figure from the Supplementary Materials of[16].

Then

$$\mathrm{GCE}^*(G_f) = \frac{E(G_f)}{E(G_0^{\mathrm{ideal}})} \tag{A.2}$$

is normalised in $[0, 1]$ and it is a monotone decreasing function w.r.t. f.

Other figures are available in the Supplementary Information of[16].

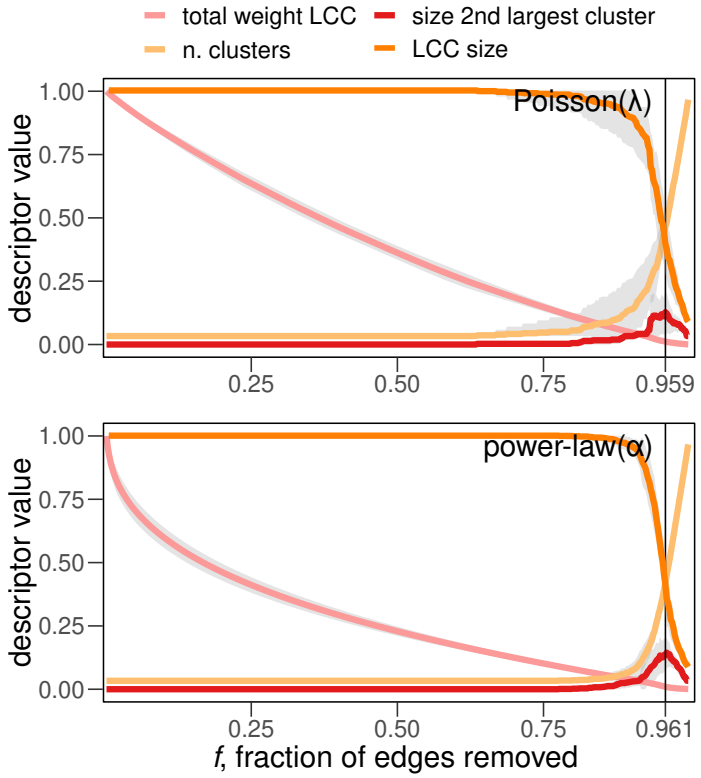

Figure A.4: Targeted bond percolation of an ensemble of synthetic networks with homogeneous and heterogeneous flows. Homogeneous flows are modelled sampling the edge weights from the Poisson distribution, while heterogeneous flows from the power-law distribution. The descriptors are the total weight of the largest connected component (LCC)—computed as the sum of edge weights in the LCC—, the size of the largest (LCC) and 2^{nd} largest clusters and the number of clusters. Lines represent the average value of the descriptors distributions over 30 networks for each model (with $\lambda = 2$ and $\alpha = 2.5$, in particular), while shaded areas the standard deviations. The vertical line indicates the critical f_c corresponding to the maximum of the size of the second largest component. Figure from the Supplementary Materials of[16].

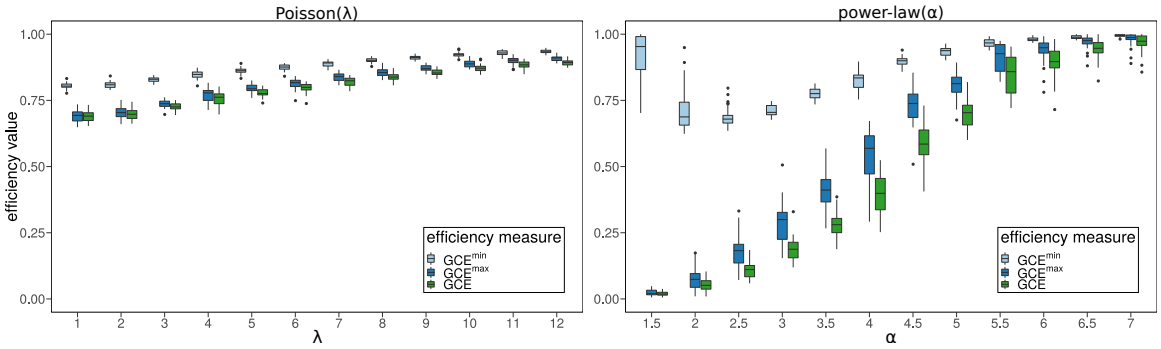

Figure A.5: Global communication efficiency (GCE) measures with variants on computation of the artificial flows. Three different GCE measures for the synthetic networks as in Fig. 2.3 with artificial flows ϕ_{ij} for each pair of nodes i, j in the networks, computed as: (GCE) total flows along shortest paths, (GCE^{min}) minimum and (GCE^{max}) maximum weight along shortest paths. All the three GCE descriptors converge to 1, but with different speeds. Additionally, GCE^{min} is not monotone and is always grater than 0.5, also for heterogeneous flows. Figure from the Supplementary Materials of[16].

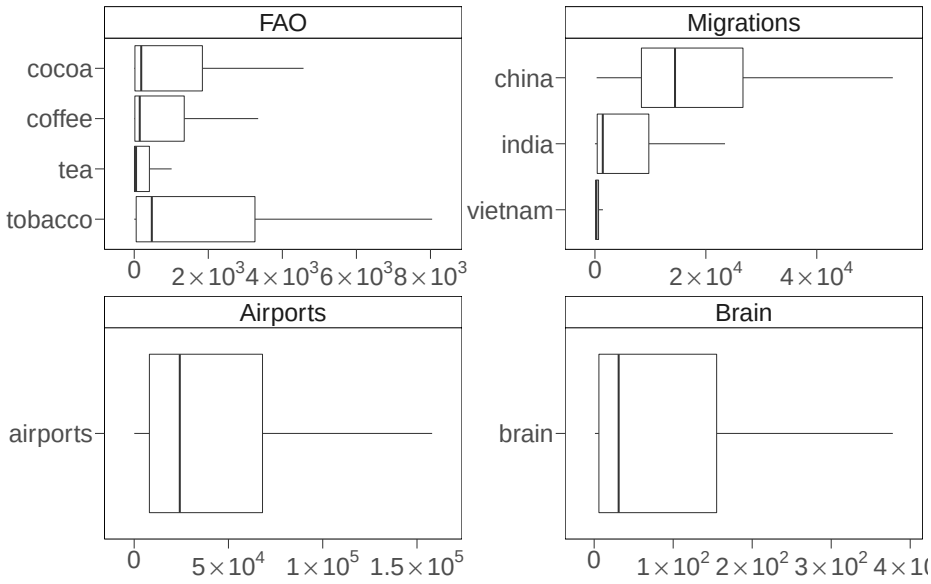

Figure A.6: Boxplots of edge weights of selected real-world networks. Heterogeneity and scale of flows vary across the datasets. From the FAO worldwide food trade network[104] we selected the layers of cocoa, coffee, tea, and tobacco. From the migration dataset[105] we selected internal migration flows inside three Asian regions: India, China and Vietnam[106]. From the worldwide air traffic network[8] we extracted the traffic in and between Europe and Africa. Finally, we consider the structural connectivity of human brain[107], quantified through diffusion tensor imaging (DTI) and fibre tractography methods. The statistics of the boxplots are the ones described ad the beginning of this appendix. Figure from the Supplementary Materials of[16].

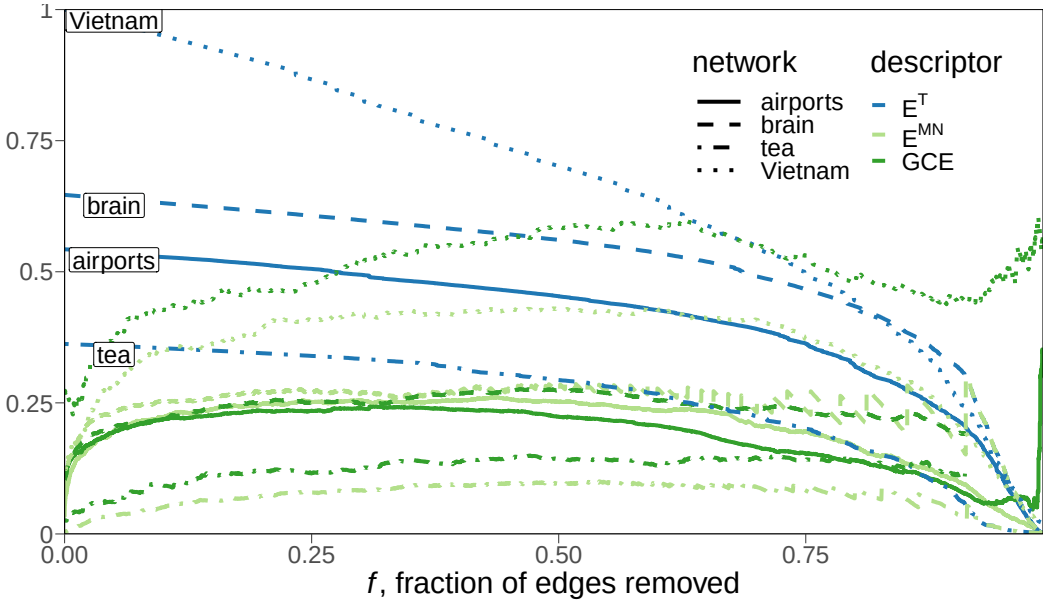

Figure A.7: Targeted bond percolation of real interconnected systems. The descriptors are the topological efficiency E^{T}, the max-normalised weighted efficiency E^{MN} and the global communication efficiency GCE of real networks where edges are removed in decreasing order according to their weight. The networks are: The tea trade network, the internal migrations in Vietnam, the air traffic in Europe and Africa and between them, and the human brain network.

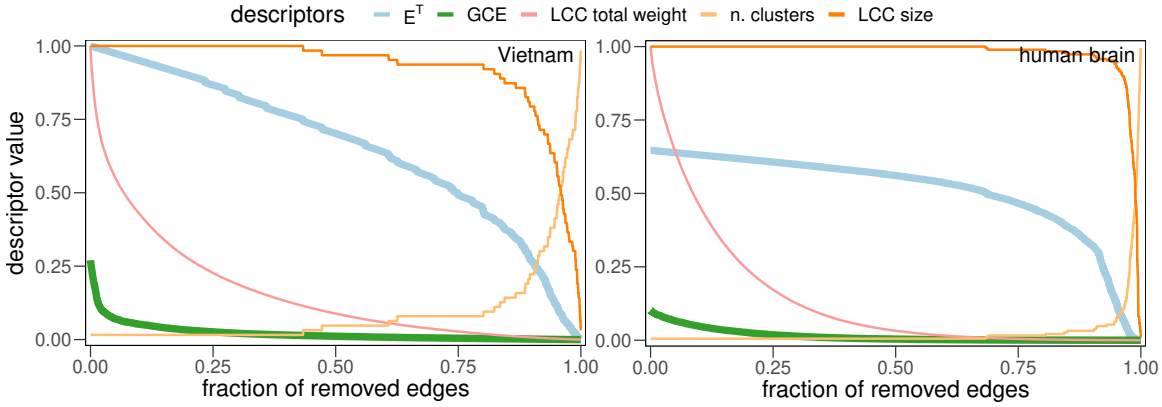

Figure A.8: Targeted bond percolation of real networks. The descriptors are: the topological efficiency E^{T} and $\mathrm{GCE}^*(G)$, i.e. the global communication efficiency (GCE) normalized on the ideal proxy G_0^{ideal} of the original undamaged network, at each step of the percolation simulation. The two real networks are the internal migrations in Vietnam[105, 106] and the human brain[107].

A.5 Functional rich-club

This appendix collects additional figures on the functional rich-club of real and model networks.

Real networks

The aggregated protein-protein interaction (PPI) networks of the Plasmodium falciparum[171, 172], the neuronal connections in the optic medulla of the Drosophila[169, 170] and the airports network[8] are described in chapter A.4. Table A.1 provides the details about two additional proteomes: the aggregated networks of interactions of the human immunodeficiency virus-1 (HIV-1) and of the Xenopus laevis, and about a coarser grained connectome of the Drosophila[74] at the level of local processing units (LPUs). All the proteome data come from the Biological General Repository for Interaction Datasets BioGRID-3.4.155[171]. From the BioGRID we obtain information about different types of interactions—e.g., physical interaction, co-localization etc.—between proteins in each organism, which is typically encoded in the different layers of a multilayer network[175], while here is aggregated into unweighted and undirected single-layer networks—we refer the interested reader to[175] for more details. Consequently, two proteins are adjacent in the resulting PPI networks if they have at least one interaction of any type. Directions and weights attached to edges have been discarded for all networks in this work.

Table A.1: Real networks. Size and average degree of some real networks. We considered different types of biological networks: the protein-protein interaction (PPI) network of the malaria parasite Plasmodium falciparum, of the HIV and of the Xenopus laevis; the connectome of the fruit fly (Drosophila melanogaster) at different scales. An instance of infrastructural network is given by the world-wide airports network. This is a the reduced version, with 965 airports out of the over four thousand airports of the complete network, proposed in[8].

	HIV PPI	X. laevis PPI	Dros. (LPUs)
n. of nodes	1180	955	49
n. of edges	1300	1095	1053
avg. degree $\langle k \rangle$	2.3	2.3	43

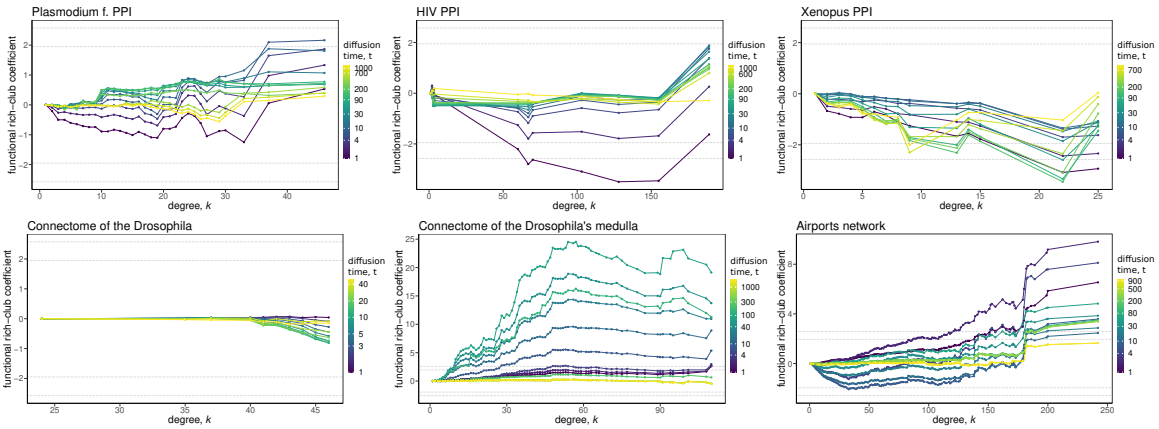

Figure A.9: Real networks and their functional rich-club organization. The top panels depict the functional rich club coefficients for three distinct protein-protein interaction (PPI) networks. In general these do not display the functional rich-club phenomenon, despite few exceptions at highest degrees and particular time scales (not persistent across time). Details and references of the networks datasets in Tabs. 4.1-A.1.

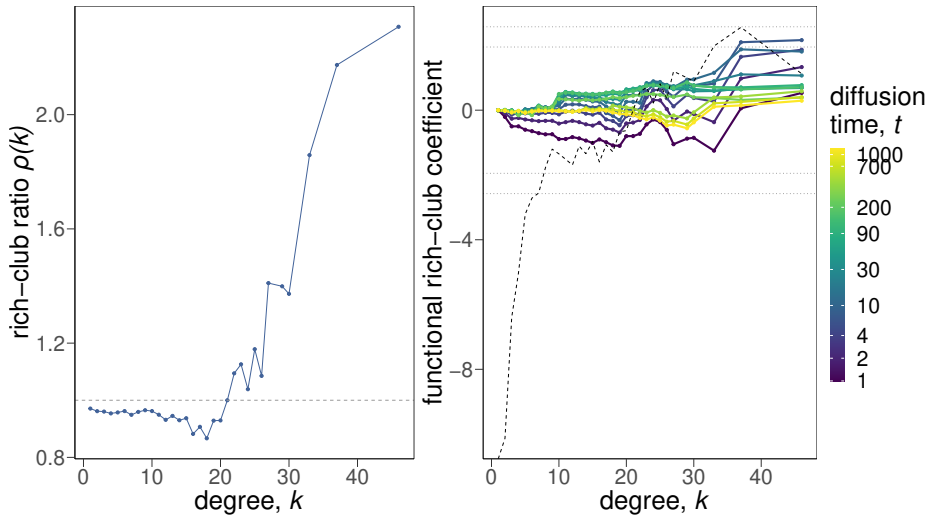

Figure A.10: Plasmodium falciparum protein-protein interaction network. Comparison of structural and functional rich-club organization.

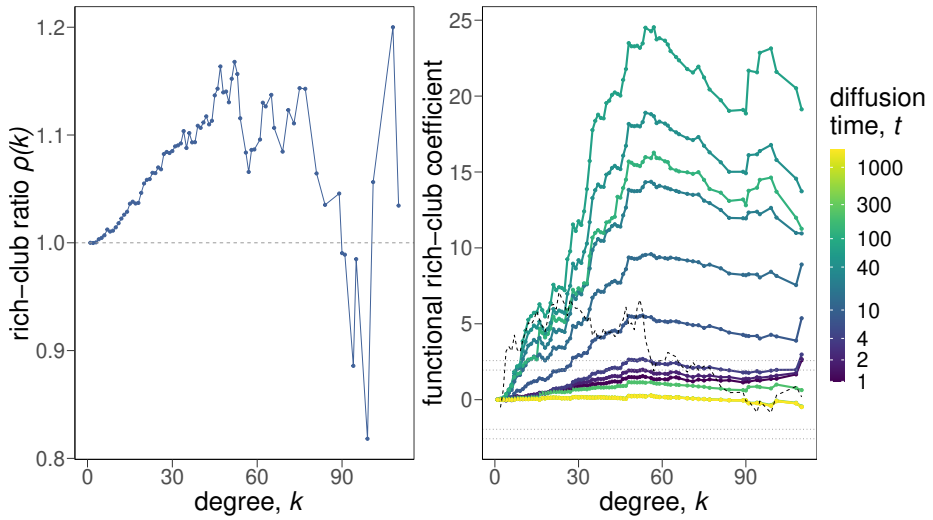

Figure A.11: Connectome of the Drosophila's medulla. Comparison of structural and functional rich-club organization.

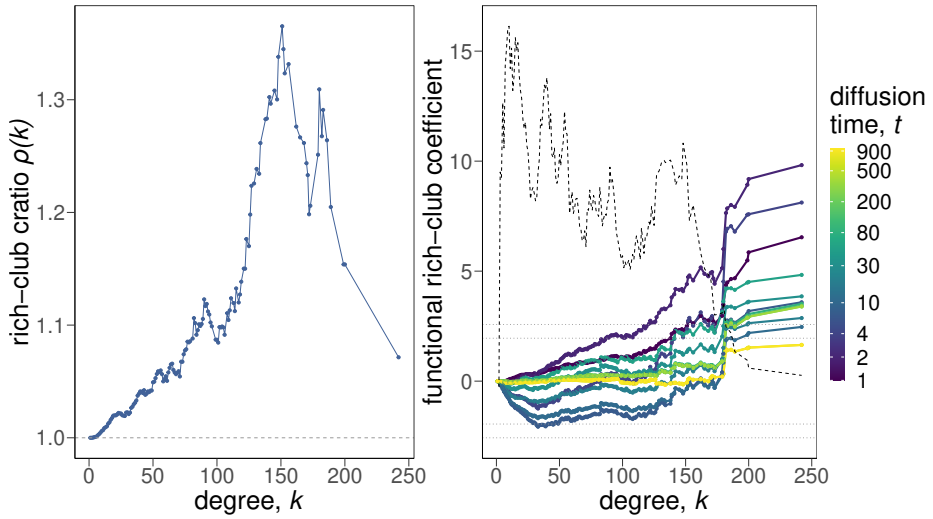

Figure A.12: World-wide airports network. Comparison of structural and functional rich-club organization.

Network models

Additionally to Barabási–Albert (BA)[102], the Erdős–Renyi (ER)[173] and the Lancichinetti–Fortunato–Radicchi (LFR)[174] models we here provide results on the Watts-Strogatz (WS) model[21], the stochastic block model (SBM)[121, 122] and on the random geometric graph (RGG)[176].

The parameters for the various network models have been appropriately chosen to yield networks with the same size ($N = 256$) and average degree $\langle k \rangle \approx 12$. In particular:

- ER: the probability of existence of an edge between an arbitrary pair of nodes is $p = \frac{12}{N}$.

- BA: the *power* parameter is set to 1, i.e., linear preferential attachment and the number of edges to add at each time step is $m = 6$.

- LFR: The network has been generated through the LFR benchmark code available at `https://github.com/eXascaleInfolab/LFR-Benchmark_UndirWeightOvp` with the following parameters: average degree $k = 12$, with minimum and maximum degree equal to 1 and 32 resp., minus exponent for degree distribution 1.5, minimum and maximum community sizes set to 16 and 64 resp., minus exponent for community size distribution equal to 1 and, finally, a community mixing parameter set to 0.1.

- WS: we start with a $\sqrt{N} \times \sqrt{N}$ square lattice (dimension 2) with each node connected to its first and second neighbours in the grid and fix the rewiring probability to 0.1.

- SBM: four communities of equal size ($\frac{N}{4} = 64$) with the following matrix of Bernoulli probabilities of inter-community and intra-communities connections (diagonal and off-diagonal entries resp.)

$$P = \begin{pmatrix} 0.98 & 0.01 & 0.01 & 0.01 \\ 0.01 & 0.98 & 0.01 & 0.01 \\ 0.01 & 0.01 & 0.98 & 0.01 \\ 0.01 & 0.01 & 0.01 & 0.98 \end{pmatrix}$$

- RGG: N points are sampled uniformly at random in the unit square (actually a torus, since we fix periodic boundary conditions) and each pair of points is connected if their (Euclidean) distances is smaller than the radius $r = 0.122$.

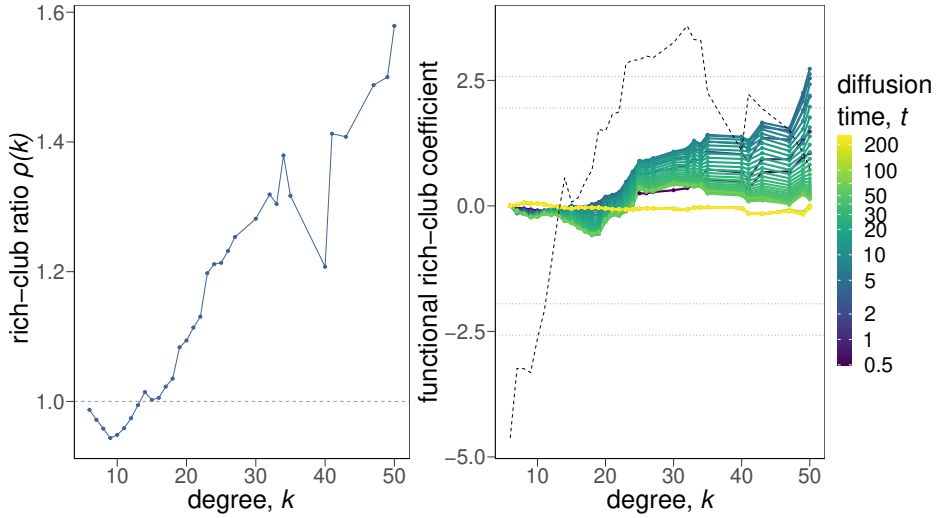

Figure A.13: BA model. Comparison of the structural and the functional rich-club coefficients.

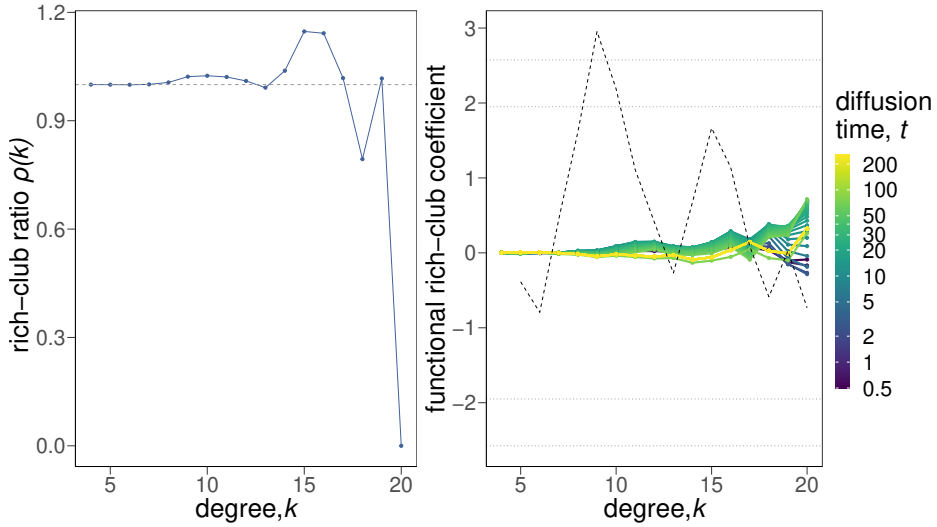

Figure A.14: ER model. Comparison of the structural and the functional rich-club coefficients.

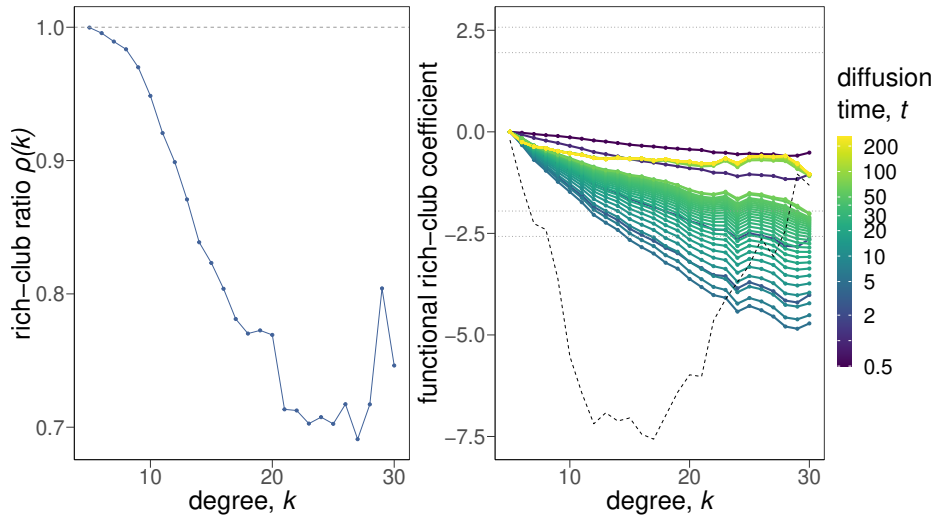

Figure A.15: LFR model. Comparison of the structural and the functional rich-club coefficients.

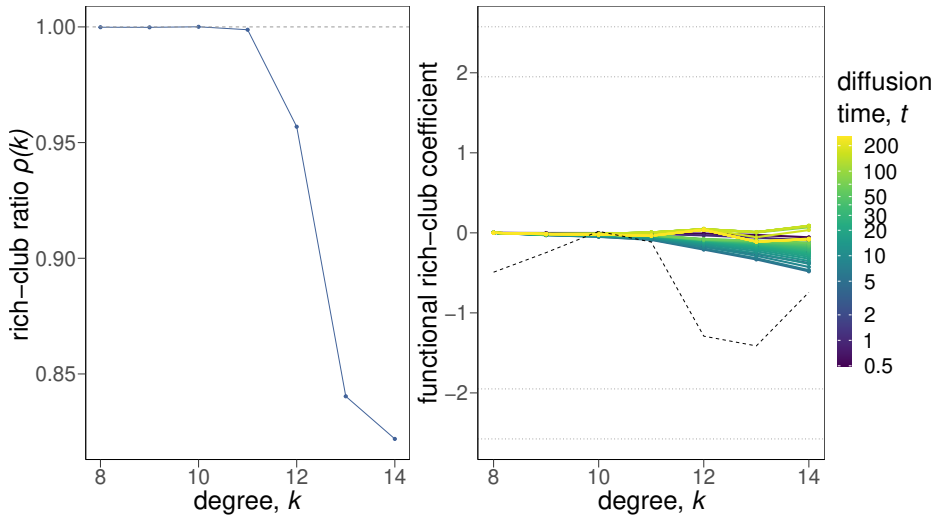

Figure A.16: WS model. Comparison of the structural and the functional rich-club coefficients.

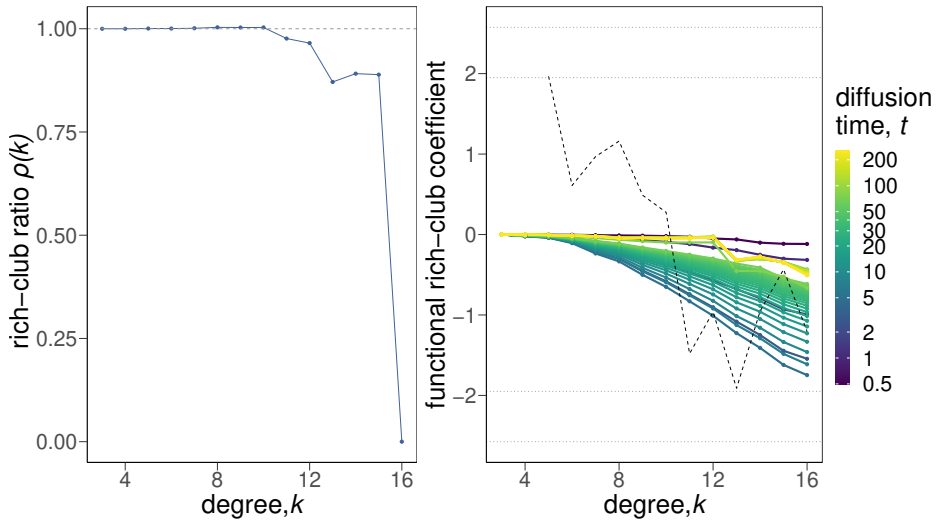

Figure A.17: SBM. Comparison of the structural and the functional rich-club coefficients.

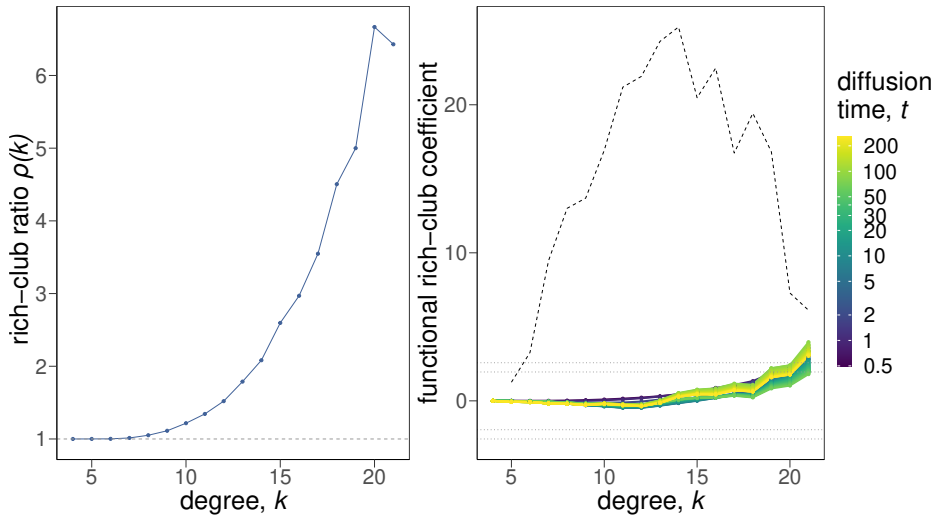

Figure A.18: RGG model. Comparison of the structural and the functional rich-club coefficients.

Appendix B

Network science terminology

In this appendix we collect some common expressions used in the literature of complex networks and systems and (partially) in this work, with the relative (tentative) definitions and examples from different fields.

Expression	**Definition and examples**
Structure	Relations and their arrangement between the units of complex systems, often modelled in terms of networks where components are nodes and their connections are the edges. Can be weighted or dyadic (topological). • In physics: The connections between states of physical systems determining the rates of transitions between them. • In biology: The network encoding connections between cells, organs or species. • In neuroscience: Gap junctions (synapses) between neurons or, at a coarser grain, fibre bundles between cortical areas. • In chemistry: The interrelations between chemical compounds in a chemical reaction network. • In social sciences: The relationships (friendship, family relations) and interactions (face-to-face, emails) between individuals in a population or social group. • In transportation systems: The network of transportation routes connecting districts, urban areas, regions, countries, or continents.
Dynamical process	The mathematical description of the change in time of a quantity or field. The evolution rule is given by a function depending on time and the variation in time is modelled with differential equations.

Expression	**Definition and examples**
	• In physics: The thermalisation protocol describing the transitions between physical states of a system. • In biology: The biochemical equation governing the spreading of chemicals or electrical signals among cells, organs or species. • In chemistry: The reaction-diffusion equations describing the behaviour of the population corresponding to a chemical reaction network. • In social sciences: The consensus dynamics between individuals in a social network, or the dynamical equations describing the spread of pathogens or news between them. • In transportation systems: The equation describing the flow of people or goods through the network of transportation routes connecting districts, urban areas, regions, countries, or continents.
Communication, also: information exchange	A process by which information (or any physical quantity) is exchanged between units of a system. More generally, it describes the effect of components of complex systems on each other, often modelled in terms of the flow of a physical quantity between pairs of nodes, governed by dynamical processes. • In physics: The exchange of particles between two physical states, induced by a thermalisation protocol or external forces. • In biology: The exchange of electro-chemical signals between two cells, organs or species. • In neuroscience: Neuro-physiological signals exchanged by neural elements through synapses or fibre bundles. • In chemistry: The influence of two chemical compounds on each other, leading to a change in their populations, in a chemical reaction network. • In social sciences: The consensus dynamics between individuals in a social network, or the dynamics describing how individuals (or computers) infect each other with a pathogen (or IT virus) or spread a piece of news. • In transportation systems: The exchange of people or goods between two nodes, through the network of transportation routes connecting districts, urban areas, regions, countries or continents.
Information flow	Emanation of a quantity or field, whose exchange between the components proxies their communications, from a source, often considered to be one of the components, into the system, through the links.

Expression	Definition and examples
	• In physics: The flow of particles from one physical state into others, induced by a thermalisation protocol or external forces. • In biology: The flow of electro-chemical signals from cells, organs or species. • In chemistry: The impact of a chemical compound on the populations of others, in a chemical reaction network. • In social sciences: The flow of pathogen or news from an individual to the rest. • In transportation systems: The flow of people or goods from one of the nodes, through the network of transportation routes connecting districts, urban areas, regions, countries, or continents.
Function	System specific tasks expected to be performed, that can involve single units or groups of them, which exchange information with each other. We stress here the common ambiguity when talking about functional organisation of a network, functional networks and function of networks, which are linked in non-trivial ways. Assuming that communication is the most basic factor for the emergence of any function in/of a system, we use the expression functional organisation to indicate how the pairwise communication shapes interactions at higher levels (e.g. between groups) and, ultimately, how these allow the system to perform its actual function(s). • In physics: The activity of a classical or quantum heat engine with certain properties like power and efficiency. • In biology: The physiological activity of a cell, organ or body. • In neuroscience: At the broadest level, cognition. • In chemistry: The characteristic behaviour of a chemical compound or groups of chemicals linked in a chemical reaction network. • In social sciences: The activity of an individual or a group of them in a society. • In transportation systems: Financial activities that depend on the flow of people or goods from one of the nodes, through the network of transportation routes connecting districts, urban areas, regions, countries or continents.

Expression	**Definition and examples**
Organisation	"The formation of spatial or temporal (or both) structures that are perceived by an external observer able to measure them in terms of information." Definition from[@Artime2022]. • In physics: Ferromagnetism, the collective behaviour due to the spin-spin interactions of electrons in a material which tend to spontaneously align at the critical temperature, effectively magnetizing the system at large scale [Artime and De Domenico, (to appear in) RSPA (2022)]. • In chemistry and biology: Turing patterns, like stripes on a fur or in a chemical reaction, and morphogenesis. • In social sciences: birds flocking, fish schooling and swarms. Social segregation in the human society. • In transportation systems: Traffic jams.

References

1. Song C, Havlin S, Makse HA (2005) Self-similarity of complex networks. Nature 433:392–395. https://doi.org/10.1038/nature03248
2. Serrano MÁ, Krioukov D, Boguñá M (2008) Self-similarity of complex networks and hidden metric spaces. Physical Review Letters 100: https://doi.org/10.1103/physrevlett.100.078701
3. Krioukov D, Papadopoulos F, Kitsak M, et al (2010) Hyperbolic geometry of complex networks. Physical Review E 82:036106. https://doi.org/10.1103/PhysRevE.82.036106
4. Boguñá M, Bonamassa I, De Domenico M, et al (2021) Network geometry. Nature Reviews Physics 3:114–135. https://doi.org/10.1038/s42254-020-00264-4
5. De Domenico M (2017) Diffusion geometry unravels the emergence of functional clusters in collective phenomena. Physical Review Letters 118:168301. https://doi.org/10.1103/PhysRevLett.118.168301
6. Estrada E, Hatano N (2008) Communicability in complex networks. Physical Review E 77: https://doi.org/10.1103/physreve.77.036111
7. Bertagnolli G, De Domenico M (2021) Diffusion geometry of multiplex and interdependent systems. Physical Review E: Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics 103:042301. https://doi.org/10.1103/PhysRevE.103.042301
8. Brockmann D, Helbing D (2013) The hidden geometry of complex, network-driven contagion phenomena. Science 342:1337–1342. https://doi.org/10.1126/science.1245200
9. Calleja N, AbdAllah A, Abad N, et al (2021) A public health research agenda for managing infodemics: Methods and results of the first WHO infodemiology conference. JMIR Infodemiology 1:e30979. https://doi.org/10.2196/30979
10. Arnaudon A, Peach RL, Barahona M (2020) Scale-dependent measure of network centrality from diffusion dynamics. Physical Review Research 2:033104. https://doi.org/10.1103/PhysRevResearch.2.033104

11. Avena-Koenigsberger A, Yan X, Kolchinsky A, et al (2019) A spectrum of routing strategies for brain networks. PLOS Computational Biology 15:e1006833. https://doi.org/10.1371/journal.pcbi.1006833
12. Schwarze AC, Porter MA (2021) Motifs for processes on networks. SIAM Journal on Applied Dynamical Systems 20:2516–2557. https://doi.org/10.1137/20m1361602
13. Artime O, Domenico MD (2022) From the origin of life to pandemics: Emergent phenomena in complex systems. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 380: https://doi.org/10.1098/rsta.2020.0410
14. Akbarzadeh M, Estrada E (2018) Communicability geometry captures traffic flows in cities. Nature Human Behaviour 2:645–652. https://doi.org/10.1038/s41562-018-0407-3
15. Bertagnolli G, Agostinelli, De Domenico M (2019) Network depth: Identifying median and contours in complex networks. Journal of Complex Networks 8: https://doi.org/10.1093/comnet/cnz041
16. Bertagnolli G, Gallotti R, Domenico MD (2021) Quantifying efficient information exchange in real network flows. Communications Physics 4: https://doi.org/10.1038/s42005-021-00612-5
17. Gallotti R, Bertagnolli G, Domenico MD (2021) Unraveling the hidden organisation of urban systems and their mobility flows. EPJ Data Sci 10:1–17. https://doi.org/10.1140/epjds/s13688-020-00258-3
18. Bollobás B (1998) Modern graph theory. Springer Science & Business Media
19. Newman MEJ (2018) Networks. Oxford university press
20. Chung FRK (1997) Spectral graph theory, 2. print. Conference board of the mathematical sciences, Providence, R.I
21. Watts DJ, Strogatz SH (1998) Collective dynamics of small-world networks. Nature 393:440–442. https://doi.org/10.1038/30918
22. Kelly FP (1979) Reversibility and stochastic networks
23. Doyle PG, Snell JL (1984) Random walks and electric networks. American Mathematical Soc.
24. Aldous D, Fill J (2002) Reversible markov chains and random walks on graphs, unfinished monograph, recompiled 2014.
25. Nadler B, Lafon S, Coifman RR, Kevrekidis IG (2005) Diffusion maps, spectral clustering and eigenfunctions of Fokker-Planck operators. In: Advances in Neural Information Processing Systems. pp 955–962
26. Coifman RR, Lafon S, Lee AB, et al (2005) Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. Proceedings of the National Academy of Sciences of the United States of America 102:7426–31. https://doi.org/10.1073/pnas.0500334102

27. Page L, Brin S, Motwani R, Winograd T (1998) A Ranking for Every Page on the Web. World Wide Web Internet And Web Information Systems Ci:1–17
28. Avena-Koenigsberger A, Misic B, Sporns O (2018) Communication dynamics in complex brain networks. Nature Reviews Neuroscience 19:17–33. https://doi.org/10.1038/nrn.2017.149
29. Holley RA, Liggett TM (1975) Ergodic theorems for weakly interacting infinite systems and the voter model. The Annals of Probability 3:643–663. https://doi.org/10.1214/aop/1176996306
30. Suchecki K, Eguíluz VM, San Miguel M (2005) Voter model dynamics in complex networks: Role of dimensionality, disorder, and degree distribution. Physical Review E: Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics 72:036132. https://doi.org/10.1103/PhysRevE.72.036132
31. Peel L, Delvenne J-C, Lambiotte R (2017) Multiscale mixing patterns in networks. Proceedings of the National Academy of Sciences of the United States of America 115:4057–4062. https://doi.org/10.1073/pnas.1713019115
32. Metropolis N, Rosenbluth AW, Rosenbluth MN, et al (1953) Equation of state calculations by fast computing machines. The Journal of Chemical Physics 21:1087–1092. https://doi.org/10.1063/1.1699114
33. Hitchcock DB (2003) A history of the metropolis–hastings algorithm. Am Stat 57:254–257. https://doi.org/10.1198/0003130032413
34. Masuda N, Porter MA, Lambiotte R (2017) Random walks and diffusion on networks. Physics Reports 716-717:1–58. https://doi.org/10.1016/j.physrep.2017.07.007
35. Kac M (1947) Random walk and the theory of brownian motion. The American Mathematical Monthly 54:369–391. https://doi.org/doi.org/10.2307/2304386
36. Bouchaud J-P, Georges A (1990) Anomalous diffusion in disordered media: Statistical mechanisms, models and physical applications. Physics reports 195:127–293. https://doi.org/10.1016/0370-1573(90)90099-N
37. Havlin S, Selinger RB, Schwartz M, et al (1988) Random multiplicative processes and transport in structures with correlated spatial disorder. Physical Review Letters 61:1438. https://doi.org/10.1103/physrevlett.61.1438
38. Havlin S, Ben-Avraham D (2002) Diffusion in disordered media. Advances in Physics 51:187–292. https://doi.org/10.1080/00018730110116353
39. Gilbert EN (1967) Minimum cost communication networks. The Bell System Technical Journal 46:2209–2227. https://doi.org/10.1002/j.1538-7305.1967.tb04250.x
40. Norris JR (1998) Markov chains. Cambridge university press
41. Lovász L (1993) Random walks on graphs: A survey. Combinatorics, Paul Erdös is Eighty 2:1–46

42. Biamonte J, Faccin M, De Domenico M (2019) Complex networks from classical to quantum. Communications Physics 2:1–10. https://doi.org/10.1038/s42005-019-0152-6
43. Böttcher L, Porter MA (2021) Classical and quantum random-walk centrality measures in multilayer networks. SIAM Journal on Applied Mathematics 81:2704–2724. https://doi.org/10.1137/20m1385998
44. Ghavasieh A, Bertagnolli G, De Domenico M (2022) Dismantling the information flow in complex interconnected systems. arXiv preprint arXiv:220209692
45. Coifman RR, Lafon S (2006) Diffusion maps. Applied and Computational Harmonic Analysis 21:5–30. https://doi.org/10.1016/j.acha.2006.04.006
46. Lambiotte R, Delvenne J-C, Barahona M (2014) Random walks, markov processes and the multiscale modular organization of complex networks. IEEE Transactions on Network Science and Engineering 1:76–90. https://doi.org/10.1109/TNSE.2015.2391998
47. Fortunato S (2010) Community detection in graphs. Physics Reports 486:75–174. https://doi.org/10.1016/j.physrep.2009.11.002
48. Rosvall M, Bergstrom CT (2008) Maps of random walks on complex networks reveal community structure. Proceedings of the National Academy of Sciences 105:1118–1123. https://doi.org/10.1073/pnas.0706851105
49. Liberti L, Lavor C, Maculan N, Mucherino A (2014) Euclidean distance geometry and applications. SIAM Review 56:3–69. https://doi.org/10.1137/120875909
50. Zhang Y, Wu H, Cheng L (2012) Some new deformation formulas about variance and covariance. In: 2012 proceedings of international conference on modelling, identification and control. pp 987–992
51. Schaub MT, Benson AR, Horn P, et al (2020) Random walks on simplicial complexes and the normalized hodge 1-laplacian. SIAM Review 62:353–391. https://doi.org/10.1137/18m1201019
52. Newman MEJ (2013) Spectral methods for community detection and graph partitioning. Physical Review E 88: https://doi.org/10.1103/physreve.88.042822
53. Belkin M, Niyogi P (2001) Laplacian eigenmaps and spectral techniques for embedding and clustering. In: Dietterich T, Becker S, Ghahramani Z (eds) Advances in neural information processing systems. MIT Press
54. Newman MEJ (2005) A measure of betweenness centrality based on random walks. Social Networks 27:39–54. https://doi.org/10.1016/j.socnet.2004.11.009
55. Klein DJ, Randić M (1993) Resistance distance. Journal of Mathematical Chemistry 12:81–95. https://doi.org/10.1007/BF01164627

56. Tetali P (1991) Random walks and the effective resistance of networks. Journal of Theoretical Probability 4:101–109. https://doi.org/10.1007/bf01046996
57. Heilbron J (2020) Geometry
58. Blanchard Ph, Dawin JR, Volchenkov D (2010) Markov chains or the game of structure and chance. The European Physical Journal Special Topics 184:1–82. https://doi.org/10.1140/epjst/e2010-01232-1
59. Gómez-Gardeñes J, Latora V (2008) Entropy rate of diffusion processes on complex networks. Physical Review E 78: https://doi.org/10.1103/physreve.78.065102
60. De Domenico M, Solé-Ribalta A, Gómez S, Arenas A (2014) Navigability of interconnected networks under random failures. Proceedings of the National Academy of Sciences 111:8351–8356. https://doi.org/10.1073/pnas.1318469111
61. Fortunato S, Hric D (2016) Community detection in networks: A user guide. Physics Reports 659:1–44. https://doi.org/10.1016/j.physrep.2016.09.002
62. Girvan M, Newman MEJ (2002) Community structure in social and biological networks. Proceedings of the national academy of sciences 99:7821–7826. https://doi.org/10.1073/pnas.122653799
63. Arenas A, Díaz-Guilera A, Pérez-Vicente CJ (2006) Synchronization reveals topological scales in complex networks. Physical Review Letters 96:114102. https://doi.org/10.1103/PhysRevLett.96.114102
64. Ronhovde P, Nussinov Z (2010) Local resolution-limit-free potts model for community detection. Physical Review E: Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics 81:046114. https://doi.org/10.1103/PhysRevE.81.046114
65. Reichardt J, Bornholdt S (2006) Statistical mechanics of community detection. Physical Review E: Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics 74:016110. https://doi.org/10.1103/PhysRevE.74.016110
66. Pons P, Latapy M (2005) Computing communities in large networks using random walks. In: Computer and information sciences - ISCIS 2005. Springer Berlin Heidelberg, pp 284–293
67. Eriksen KA, Simonsen I, Maslov S, Sneppen K (2003) Modularity and extreme edges of the internet. Physical Review Lettetters 90:148701. https://doi.org/10.1103/PhysRevLett.90.148701
68. Donetti L, Muñoz MA (2004) Detecting network communities: A new systematic and efficient algorithm. Journal of Statistical Mechanics: Theory and Experiment 2004:P10012. https://doi.org/10.1088/1742-5468/2004/10/p10012

69. Schaub MT, Delvenne J-C, Lambiotte R, Barahona M (2019) Multiscale dynamical embeddings of complex networks. Physical Review E: Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics 99:062308. https://doi.org/10.1103/PhysRevE.99.062308
70. Perozzi B, Al-Rfou R, Skiena S (2014) DeepWalk: Online learning of social representations. In: Proceedings of the 20th ACM SIGKDD international conference on knowledge discovery and data mining. Association for Computing Machinery, New York, NY, USA, pp 701–710
71. Tukey JW (1975) Mathematics and the picturing of data. In: Proceedings of the international congress of mathematicians, vancouver, 1975. pp 523–531
72. Liu RY, Parelius JM, Singh K, et al (1999) Multivariate analysis by data depth: Descriptive statistics, graphics and inference,(with discussion and a rejoinder by liu and singh). The annals of statistics 27:783–858
73. Zuo Y, Serfling R (2000) General notions of statistical depth function. Annals of statistics 461–482
74. Shih C-T, Sporns O, Yuan S-L, et al (2015) Connectomics-based analysis of information flow in the drosophila brain. Current Biology 25:1249–1258
75. Freeman LC (1978) Centrality in social networks conceptual clarification. Social Networks 1:215–239. https://doi.org/10.1016/0378-8733(78)90021-7
76. Olfati-Saber R, Murray RM (2004) Consensus problems in networks of agents with switching topology and time-delays. IEEE Transactions on Automatic Control 49:1520–1533. https://doi.org/10.1109/TAC.2004.834113
77. Blanchard P, Volchenkov D (2011) Random Walks and Diffusions on Graphs and Databases: An Introduction. Springer Science & Business Media
78. Kondor RI, Lafferty J (2002) Diffusion kernels on graphs and other discrete structures. In: Proceedings of the 19th international conference on machine learning. pp 315–322
79. Chung F (2007) The heat kernel as the pagerank of a graph. Proceedings of the National Academy of Sciences 104:19735–19740. https://doi.org/10.1073/pnas.0708838104
80. Bertagnolli G, De Domenico M (2021) Diffudist: Diffusion distance for complex networks
81. Latora V, Marchiori M (2001) Efficient behavior of small-world networks. Physical Review Letters 87:198701. https://doi.org/10.1103/physrevlett.87.198701
82. Newman MEJ (2003) The structure and function of complex networks. SIAM Review 45:167–256. https://doi.org/10.1137/S003614450342480
83. Latora V, Marchiori M (2003) Economic small-world behavior in weighted networks. The European Physical Journal B-Condensed Matter and Complex Systems 32:249–263. https://doi.org/10.1140/epjb/e2003-00095-5

84. Barrat A, Barthelemy M, Pastor-Satorras R, Vespignani A (2004) The architecture of complex weighted networks. Proceedings of the National Academy of Sciences of the United States of America 101:3747–3752. https://doi.org/10.1073/pnas.0400087101
85. Granovetter MS (1973) The Strength of Weak Ties. American Journal of Sociology 78:1360–1380. https://doi.org/10.1093/oso/9780195159509.003.0010
86. Dijkstra EW (1959) A note on two problems in connexion with graphs. Numerische mathematik 1:269–271. https://doi.org/10.1007/bf01386390
87. Floyd RW (1962) Algorithm 97: Shortest path. Communications of the ACM 5:345. https://doi.org/10.1145/367766.368168
88. Newman MEJ (2001) Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. Physical Review E: Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics 64:016132. https://doi.org/10.1103/physreve.64.016132
89. Brandes U (2001) A faster algorithm for betweenness centrality. Journal of Mathematical Sociology 25:163–177. https://doi.org/10.1080/0022250x.2001.9990249
90. Opsahl T, Agneessens F, Skvoretz J (2010) Node centrality in weighted networks: Generalizing degree and shortest paths. Soc Netw 32:245–251. https://doi.org/10.1016/j.socnet.2010.03.006
91. Rubinov M, Sporns O (2010) Complex network measures of brain connectivity: Uses and interpretations. Neuroimage 52:1059–1069. https://doi.org/10.1016/j.neuroimage.2009.10.003
92. Bullmore ET, Bassett DS (2011) Brain graphs: Graphical models of the human brain connectome. Annu Rev Clin Psycho 7:113–140. https://doi.org/10.1146/annurev-clinpsy-040510-143934
93. Watson CG (2019) brainGraph: Graph theory analysis of brain MRI data
94. Bellingeri M, Bevacqua D, Scotognella F, Cassi D (2019) The heterogeneity in link weights may decrease the robustness of real-world complex weighted networks. Scientific Reports 9:1–13. https://doi.org/10.1038/s41598-019-47119-2
95. Achard S, Bullmore E (2007) Efficiency and cost of economical brain functional networks. PLOS Computational Biology 3:1–10. https://doi.org/10.1371/journal.pcbi.0030017
96. Bertagnolli G (2021) Intsegration: Integration and segregation of complex networks
97. Bertagnolli G Repository for data and code related to [Commun Phys 4, 1–10 (2021)]

98. Clauset A, Shalizi CR, Newman MEJ (2009) Power-law distributions in empirical data. SIAM Review 51:661–703. https://doi.org/10.1137/070710111
99. Latora V, Marchiori M (2005) Vulnerability and protection of infrastructure networks. Physical Review E: Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics 71:015103. https://doi.org/10.1103/PhysRevE.71.015103
100. Viles W, Ginestet CE, Tang A, et al (2016) Percolation under noise: Detecting explosive percolation using the second-largest component. Physical Review E: Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics 93:052301. https://doi.org/10.1103/PhysRevE.93.052301
101. Silva CR da, Lyra ML, Viswanathan GM (2002) Largest and second largest cluster statistics at the percolation threshold of hypercubic lattices. Physical Review E: Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics 66:056107. https://doi.org/10.1103/PhysRevE.66.056107
102. Barabási A-L, Albert R (1999) Emergence of scaling in random networks. Science 286:509–512. https://doi.org/10.1126/science.286.5439.509
103. Motter AE, Zhou CS, Kurths J (2005) Enhancing complex-network synchronization. Europhysics Letters 69:334–340. https://doi.org/10.1209/epl/i2004-10365-4
104. De Domenico M, Solé-Ribalta A, Omodei E, et al (2015) Ranking in interconnected multilayer networks reveals versatile nodes. Nature Communications 6:6868. https://doi.org/10.1038/ncomms7868
105. WorldPop
106. Sorichetta A, Bird TJ, Ruktanonchai NW, et al (2016) Mapping internal connectivity through human migration in malaria endemic countries. Scientific Data 3:160066. https://doi.org/10.1038/sdata.2016.66
107. Brown JA, Rudie JD, Bandrowski A, et al (2012) The UCLA multimodal connectivity database: A web-based platform for brain connectivity matrix sharing and analysis. Frontiers in Neuroinformatics 6:28. https://doi.org/10.3389/fninf.2012.00028
108. Bassett DS, Sporns O (2017) Network neuroscience. Nature Neuroscience 20:353–364. https://doi.org/10.1038/nn.4502
109. Goñi J, Avena-Koenigsberger A, Mendizabal NV de, et al (2013) Exploring the Morphospace of Communication Efficiency in Complex Networks. PLOS ONE 8:e58070. https://doi.org/10.1371/journal.pone.0058070
110. Berlingerio M, Coscia M, Giannotti F, et al (2011) Foundations of multidimensional network analysis. In: 2011 international conference on advances in social networks analysis and mining. pp 485–489
111. Kivelä M, Arenas A, Barthélemy M, et al (2014) Multilayer networks. Journal of Complex Networks 2:203–271. https://doi.org/10.1093/comnet/cnu016

112. De Domenico M, Solé-Ribalta A, Cozzo E, et al (2013) Mathematical formulation of multilayer networks. Physical Review X 3:041022. https://doi.org/10.1103/PhysRevX.3.041022
113. Brin S, Page L (1998) The anatomy of a large-scale hypertextual web search engine. Computer Networks and ISDN Systems 30:107–117. https://doi.org/10.1016/s0169-7552(98)00110-x
114. De Domenico M, Lancichinetti A, Arenas A, Rosvall M (2015) Identifying modular flows on multilayer networks reveals highly overlapping organization in interconnected systems. Physical Review X 5:011027. https://doi.org/10.1103/PhysRevX.5.011027
115. Gomez S, Diaz-Guilera A, Gomez-Gardenes J, et al (2013) Diffusion dynamics on multiplex networks. Physical Review Letters 110:028701. https://doi.org/10.1103/physrevlett.110.028701
116. Kolda TG, Bader BW (2009) Tensor decompositions and applications. SIAM Review 51:455–500. https://doi.org/10.1137/07070111x
117. Battiston F, Nicosia V, Latora V (2016) Efficient exploration of multiplex networks. New Journal of Physics 18: https://doi.org/10.1088/1367-2630/18/4/043035
118. Ghavasieh A, Domenico MD (2020) Enhancing transport properties in interconnected systems without altering their structure. Physical Review Research 2: https://doi.org/10.1103/PhysRevResearch.2.013155
119. Burda Z, Duda J, Luck J-M, Waclaw B (2009) Localization of the maximal entropy random walk. Physical Review Letters 102:160602. https://doi.org/10.1103/PhysRevLett.102.160602
120. Sinatra R, Gómez-Gardeñes J, Lambiotte R, et al (2011) Maximal-entropy random walks in complex networks with limited information. Physical Review E 83: https://doi.org/10.1103/physreve.83.030103
121. Holland PW, Laskey KB, Leinhardt S (1983) Stochastic blockmodels: First steps. Social Networks 5:109–137. https://doi.org/10.1016/0378-8733(83)90021-7
122. Faust K, Wasserman S (1992) Blockmodels: Interpretation and evaluation. Social Networks 14:5–61. https://doi.org/10.1016/0378-8733(92)90013-w
123. Team RC (2019) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria
124. Edler D, Bohlin L, Rosvall M (2017) Mapping higher-order network flows in memory and multilayer networks with infomap. Algorithms 10:112
125. Perotti JI, Almeira N, Saracco F (2020) Towards a generalization of information theory for hierarchical partitions. Physical Review E 101:062148. https://doi.org/10.1103/PhysRevE.101.062148

126. Roberts N, Everton SF (2011) Terrorist data: Noordin top terrorist network (subset).[machine-readable data file]
127. Battiston F, Nicosia V, Latora V (2014) Structural measures for multiplex networks. Physical Review E 89:032804. https://doi.org/10.1103/PhysRevE.89.032804
128. Leeuw J de, Mair P (2009) Multidimensional Scaling Using Majorization: SMACOF in R. Journal of Statistical Software 31:1–30. https://doi.org/10.18637/jss.v031.i03
129. Mantel N (1967) The detection of disease clustering and a generalized regression approach. Cancer Research 27:209–220
130. Legendre P, Legendre L (2012) Chapter 10 - interpretation of ecological structures. In: Legendre P, Legendre L (eds) Numerical ecology. Elsevier, Oxford, UK, pp 521–624
131. Cox TF, Cox MAA (2000) Multidimensional scaling. Chapman; hall/CRC
132. Zhang X, Boccaletti S, Guan S, Liu Z (2015) Explosive synchronization in adaptive and multilayer networks. Physical review letters 114:038701
133. Arruda GF de, Petri G, Moreno Y (2020) Social contagion models on hypergraphs. Physical Review Research 2:023032
134. Papadopoulos F, Kitsak M, Serrano MÁ, et al (2012) Popularity versus similarity in growing networks. Nature 489:537–540. https://doi.org/10.1038/nature11459
135. Bianconi G, Rahmede C (2017) Emergent hyperbolic network geometry. Scientific Reports 7:41974. https://doi.org/10.1038/srep41974
136. Kleineberg K-K, Boguñá M, Serrano MÁ, Papadopoulos F (2016) Hidden geometric correlations in real multiplex networks. Nature Physics 12:1076–1081. https://doi.org/10.1038/nphys3812
137. Battiston F, Cencetti G, Iacopini I, et al (2020) Networks beyond pairwise interactions: Structure and dynamics. Physics Reports. https://doi.org/10.1016/j.physrep.2020.05.004
138. Newman MEJ (2002) Assortative mixing in networks. Physical Review Letters 89:208701. https://doi.org/10.1103/PhysRevLett.89.208701
139. Zhou S, Mondragón RJ (2004) The rich-club phenomenon in the internet topology. IEEE Communications Letters 8:180–182. https://doi.org/10.1109/LCOMM.2004.823426
140. Ball G, Aljabar P, Zebari S, et al (2014) Rich-club organization of the newborn human brain. Proceedings of the National Academy of Sciences of the United States of America 111:7456–7461. https://doi.org/10.1073/pnas.1324118111
141. Betzel RF, Bassett DS (2017) Multi-scale brain networks. Neuroimage 160:73–83. https://doi.org/10.1016/j.neuroimage.2016.11.006

142. Mišić B, Betzel RF, Reus MA de, et al (2016) Network-level structure-function relationships in human neocortex. Cerebral Cortex 26:3285–3296. https://doi.org/10.1093/cercor/bhw089
143. Harriger L, Heuvel MP van den, Sporns O (2012) Rich club organization of macaque cerebral cortex and its role in network communication. PLoS ONE 7:e46497. https://doi.org/10.1371/journal.pone.0046497
144. Senden M, Deco G, Reus MA de, et al (2014) Rich club organization supports a diverse set of functional network configurations. NeuroImage 96:174–182. https://doi.org/10.1016/j.neuroimage.2014.03.066
145. Van Den Heuvel MP, Sporns O (2011) Rich-club organization of the human connectome. Journal of Neuroscience 31:15775–15786. https://doi.org/10.1523/JNEUROSCI.3539-11.2011
146. Moretti P, Hütt M-T (2020) Link-usage asymmetry and collective patterns emerging from rich-club organization of complex networks. Proceedings of the National Academy of Sciences 117:18332–18340. https://doi.org/10.1073/pnas.1919785117
147. Arnatkeviciute A, Fulcher BD, Oldham S, et al (2021) Genetic influences on hub connectivity of the human connectome. Nature Communications 12: https://doi.org/10.1038/s41467-021-24306-2
148. Heuvel MP van den, Sporns O, Collin G, et al (2013) Abnormal rich club organization and functional brain dynamics in schizophrenia. JAMA Psychiatry 70:783–792
149. Daianu M, Mezher A, Mendez MF, et al (2015) Disrupted rich club network in behavioral variant frontotemporal dementia and early-onset alzheimer's disease. Human Brain Mapping 37:868–883. https://doi.org/10.1002/hbm.23069
150. Daianu M, Jahanshad N, Nir TM, et al (2015) Rich club analysis in the alzheimer's disease connectome reveals a relatively undisturbed structural core network. Human Brain Mapping 36:3087–3103. https://doi.org/10.1002/hbm.22830
151. Gollo LL, Zalesky A, Hutchison RM, et al (2015) Dwelling quietly in the rich club: Brain network determinants of slow cortical fluctuations. Philosophical Transactions of the Royal Society B: Biological Sciences 370:20140165. https://doi.org/10.1098/rstb.2014.0165
152. Li R, Liao W, Li Y, et al (2016) Disrupted structural and functional rich club organization of the brain connectome in patients with generalized tonic-clonic seizure. Human Brain Mapping 37:4487–4499. https://doi.org/10.1002/hbm.23323

153. Liu X, He C, Fan D, et al (2021) Disrupted rich-club network organization and individualized identification of patients with major depressive disorder. Progress in Neuro-Psychopharmacology and Biological Psychiatry 108:110074. https://doi.org/10.1016/j.pnpbp.2020.110074
154. Opsahl T, Colizza V, Panzarasa P, Ramasco JJ (2008) Prominence and control: The weighted rich-club effect. Physical Review Letters 101:168702. https://doi.org/10.1103/PhysRevLett.101.168702
155. Ma A, Mondragón RJ, Latora V (2015) Anatomy of funded research in science. Proceedings of the National Academy of Sciences 112:14760–14765. https://doi.org/10.1073/pnas.1513651112
156. Szell M, Sinatra R (2015) Research funding goes to rich clubs. Proceedings of the National Academy of Sciences 112:14749–14750. https://doi.org/10.1073/pnas.1520118112
157. Colizza V, Flammini A, Serrano MA, Vespignani A (2006) Detecting rich-club ordering in complex networks. Nature Physics 2:110–115. https://doi.org/10.1038/nphys209
158. Pedreschi N, Battaglia D, Barrat A (2021) The temporal rich club phenomenon
159. Jiang Z-Q, Zhou W-X (2008) Statistical significance of the rich-club phenomenon in complex networks. New Journal of Physics 10:043002. https://doi.org/10.1088/1367-2630/10/4/043002
160. Zhou S, Mondragón RJ (2007) Structural constraints in complex networks. New Journal of Physics 9: https://doi.org/10.1088/1367-2630/9/6/173
161. Cinelli M, Ferraro G, Iovanella A (2018) Rich-club ordering and the dyadic effect: Two interrelated phenomena. Physica A: Statistical Mechanics and its Applications 490:808–818. https://doi.org/10.1016/j.physa.2017.08.122
162. Serrano MÁ (2008) Rich-club vs rich-multipolarization phenomena in weighted networks. Physical Review E 78: https://doi.org/10.1103/PhysRevE.78.026101
163. Cinelli M (2019) Generalized rich-club ordering in networks. Journal of Complex Networks 7:702–719. https://doi.org/10.1093/comnet/cnz002
164. Zlatic V, Bianconi G, Diaz-Guilera A, et al (2009) On the rich-club effect in dense and weighted networks. The European Physical Journal B 67:271–275. https://doi.org/10.1140/epjb/e2009-00007-9
165. McAuley JJ, Fontoura Costa L da, Caetano TS (2007) Rich-club phenomenon across complex network hierarchies. Applied Physics Letters 91:084103. https://doi.org/10.1063/1.2773951
166. Molloy M, Reed B (1995) A critical point for random graphs with a given degree sequence. Random Structures & Algorithms 6:161–180. https://doi.org/10.1002/rsa.3240060204

167. Bohlin L, Edler D, Lancichinetti A, Rosvall M (2014) Community detection and visualization of networks with the map equation framework. In: Ding Y, Rousseau R, Wolfram D (eds) Measuring scholarly impact: Methods and practice. Springer International Publishing, Cham, pp 3–34
168. Edler D, Rosvall M (2013) The MapEquation software package
169. Takemura S, Bharioke A, Lu Z, et al (2013) A visual motion detection circuit suggested by Drosophila connectomics. Nature 500:175–181. https://doi.org/10.1038/nature12450
170. Vogelstein JT, Perlman E, Falk B, et al (2018) A community-developed open-source computational ecosystem for big neuro data. Nature Methods 15:846–847. https://doi.org/10.1038/s41592-018-0181-1
171. Chatr-aryamontri A, Oughtred R, Boucher L, et al (2016) The BioGRID interaction database: 2017 update. Nucleic Acids Research 45:D369–D379. https://doi.org/10.1093/nar/gkw1102
172. Stark C (2006) BioGRID: A general repository for interaction datasets. Nucleic Acids Research 34:D535–D539. https://doi.org/10.1093/nar/gkj109
173. Erdős P, Rényi A (1959) On random graphs. Publicationes Mathematicae Debrecen 6:290–297. https://doi.org/10.2307/1999405
174. Lancichinetti A, Fortunato S, Radicchi F (2008) Benchmark graphs for testing community detection algorithms. Physical Review E: Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics 78:046110. https://doi.org/10.1103/PhysRevE.78.046110
175. De Domenico M, Nicosia V, Arenas A, Latora V (2015) Structural reducibility of multilayer networks. Nature Communications 6: https://doi.org/10.1038/ncomms7864
176. Penrose M (2003) Random geometric graphs. Oxford University Press