Statistical physics of network structure and information dynamics

A Ghavasieh1,2,☆ and M De Domenico2,3,☆

1 Department of Physics, University of Trento, Via Sommarive, 14, 38123 Povo (TN), Italy
2 ComiNe Lab, Fondazione Bruno Kessler, Via Sommarive 18, 38123 Povo (TN), Italy
3 Department of Physics and Astronomy “Galileo Galilei”, University of Padova, Padova, Italy

☆ Author to whom any correspondence should be addressed.

E-mail: manlio.dedomenico@unipd.it

Keywords: network entropy, network density matrix, information theory, statistical physics, information dynamics

Abstract
In the last two decades, network science has proven to be an invaluable tool for the analysis of empirical systems across a wide spectrum of disciplines, with applications to data structures admitting a representation in terms of complex networks. On the one hand, especially in the last decade, an increasing number of applications based on geometric deep learning have been developed to exploit, at the same time, the rich information content of a complex network and the learning power of deep architectures, highlighting the potential of techniques at the edge between applied math and computer science. On the other hand, studies at the edge of network science and quantum physics are gaining increasing attention, e.g., because of the potential applications to quantum networks for communications, such as the quantum Internet. In this work, we briefly review a novel framework grounded on statistical physics and techniques inspired by quantum statistical mechanics which have been successfully used for the analysis of a variety of complex systems. The advantage of this framework is that it allows one to define a set of information-theoretic tools which find widely used counterparts in machine learning and quantum information science, while providing a grounded physical interpretation in terms of a statistical field theory of information dynamics. We discuss the most salient theoretical features of this framework and selected applications to protein–protein interaction networks, neuronal systems, social and transportation networks, as well as potential novel applications for quantum network science and machine learning.

1. Introduction

Empirical systems exhibit peculiar features, from the abundance of topological shortcuts which reduce the characteristic geodesic length of the network and favor quick and efficient information exchange—a phenomenon known as small-world [1]—to a high degree of heterogeneity in the connectivity distribution [2]. Application of network science for modeling and analysis of real-world systems are ubiquitous, ranging from cell metabolism [3] to the human brain [4, 5], from social and socio-technical systems [6, 7] to large-scale communication networks such as the Internet [8], allowing to gain insights about life and disease [9–11], human dynamics [12], the fragility of ecosystems [13–16] and collective behavior [17, 18], to mention a few representative examples. The success of network science builds on the development of theoretical and computational tools which are able to capture a variety of features of complex and adaptive systems. The structure and dynamics of both static and temporal networks [19], for instance in terms of their latent geometry [20–24] (see [25] for a recent review), mesoscale organization [26, 27] (see [28, 29] for a thorough review), multilayer, interdependent [30–35] (see [36–39] for a thorough review) and, in general, higher-order interactions [40, 41] (see [42, 43] for a thorough review), has been extensively investigated at the edge of statistical physics, nonlinear dynamics and computer science. Besides classical systems, network science has found a broad basin of applications in the realm of quantum physics and technologies, since quantum networks open the door to
quantum nonlocality [44], distributed quantum computation [45], quantum sensing and large-scale quantum
communication such as the quantum Internet [46, 47], thanks to their ability to distribute and store entangle-
ment among remote nodes [48, 49]. Interestingly, classical and quantum random networks exhibit different

critical properties, the most striking one being in the critical probability for the appearance of subgraphs, like
cycles and trees, and its scaling with system size [50].

The importance of dealing with relational and structural data in real-world applications has been also
recognized by the machine learning community, with the flourishing, in the last years, of techniques designed
to generalize deep learning methods to non-Euclidean data, e.g., graphs and manifolds. Such approaches have
been developed [51, 52], allowing to operate directly on graphs with efficient and scalable algorithms [53],
thus boosting the number of applications across disciplines, ranging from web-scale recommender systems
[54] to chemical reaction networks [55] and resilience of interconnected infrastructures [56].

A common feature of these approaches and their development is how they deal with information to gain
insights about a system, a topic that will be the subject of section 6, where we overview the different attempts
to build an information theory of complex networks. The reminder of this work is organized as follows. In
section 3 we introduce the above mathematical framework in terms of a statistical field theory of complex
information dynamics, describing the network density matrix in terms of stream operators guiding the flow
of information. We discuss the corresponding partition function and its physical interpretation as dynamical
trapping, measuring the tendency of the structure to hinder the flow of information. In section 4 we introduce
some fundamental quantities, such as information and relative entropies, as well as measures which allow one
to quantify the functional diversity of units, as a consequence of availability of diverse diffusion pathways, and
the distance between pairs of complex networks. In section 5 we will describe applications to a variety of empir-
cical systems across scales, from cells to societies, mentioning a range of topics explored within the frameworks,
from functional reducibility of multilayer systems to robustness and mesoscale organization analysis. Finally,
in section 7 we discuss about the increasingly central role that the information is playing in physics, machine
learning and network science, providing an integrated vision and highlighting how the techniques presented
in this work might provide the ground for a unifying framework, mentioning current challenges and future
opportunities.

2. Information content and processing in complex networks

In fact, physical systems—them being biological networks, quantum networks or large-scale infrastruc-
tures—can be characterized in terms of their information capacity and information processing [57], which
can be mathematically mapped as the changes in time from an input state to an output state. While clas-
sical information theory for the analysis of unstructured data has been introduced almost one century ago,
and developed along decades even with quantum generalizations, an information theory for structured data
(figure 1) is still missing. At this point, it is worth differentiating the major challenges to this aim: (i) quanti-
fying the information content or the complexity of a network, a problem related to how information is stored
within a system state and its changes, and (ii) understanding how networked units process individually and
collectively information to achieve a desired function. While a deeper discussion about what is information in
general [58] will be presented in section 7, here we briefly discuss the existing attempts towards both directions.

On the one hand, the information content of networks can be estimated by considering a set of descriptors
(e.g., degrees, paths, so forth and so on) that can be used as structural constraints that characterize specific net-
work ensembles obtained from a max-entropy approach [60–62]: the entropy, calculated as the Gibbs entropy
(i.e., the logarithm of the number of graphs for a microcanonical ensemble) or the Shannon entropy (for a
canonical ensemble), provides a measure of disorder or surprise about the ensemble itself. On the other hand, a
widely used approach has been to use those descriptors, estimate their joint probability and calculate the Shan-
non entropy of the resulting distribution: the choice of different descriptors characterizes different methods
[63–66] (see [67] for a review), which have been used to unravel the existence of constraints operating on the
possible Universe of complex networks [59]. One drawback of the first approach is that it does not explicitly
account for information dynamics and its interplay with the network structure, therefore it cannot be used
to gain insights about node–node communications, information processing and system’s function. Moreover,
it does not allow to explain the emergence of heterogeneity in complex networks, a problem recently solved
by introducing a novel classical network ensemble and soft constraints [68]. One drawback of the second
approach is that it can only rely on a finite number of descriptors extracted from the network, thus inevitably
accounting for limited information, and estimations depend on the specific choice of those descriptors, thus
complicating the comparison of results across different studies or leading to conflicting outcomes. It is worth
mentioning that despite the drawback, classical network information theory approach has been successfully
applied to detect patterns in empirical networks, reconstruct network properties from partial information,
and sample networks with given properties [69]. To overcome the problem of limited information, it has been
Figure 1. Qualitative space identified by three ubiquitous features of networks: randomness, heterogeneity and modularity. These three dimensions, along with several other ones, are responsible for the complexity of real-world systems that statistical physics and information theory of networks aim to capture. Reprinted by permission from Springer Nature Customer Service Centre GmbH: Springer. [59] (c) 2004.

proposed to use matrix functions encoding, explicitly, the whole connectivity information. An early attempt [70] was based on defining the network counterpart of a density matrix, an operator widely used in quantum statistical physics because it allows to encode the quantum state of a system, calculate the probability and the expected value of quantum measurements. It is worth noting that while such approaches are inspired by the mathematical formulation of quantum mechanics, they are often proposed to analyse classical networks. This type of density matrix, assumed to be proportional to the combinatorial Laplacian of a network, and the corresponding von Neumann entropy, have been used in a variety of studies [71–75]. However, while such a density matrix satisfies the required mathematical properties, it lacks physical interpretation and the corresponding information entropy is does not satisfy theoretical requirements such as sub-additivity.

Concerning information processing, it is widely accepted that units in real-world systems exchange information quite efficiently without necessarily needing global intelligence, i.e., information about the full network connectivity. Therefore, there are local mechanisms at work that, together with the peculiar structural characteristics, allow one to navigate the network and information to flow efficiently [4, 8, 76]. A standard way is to use specific classes of dynamics on the top of the network and exploit them to capture specific aspects of system’s function and organization. One example is given by random walk dynamics, used to build maps of how information flows through the system: the information-theoretic content of these maps, for instance in terms of the description length of walk trajectories, can be optimized to identify the mesoscale organization of the network, i.e., the partition of system’s units in groups or modules which requires the minimum number of bits to be described. Therefore, this class of approaches exploits the presence of regularities to build a compressed description of the system [77, 78]. Another approach, not necessarily limited to random walk dynamics, is to use a perturbative formalism to track the contribution of each node and path to the overall flow of information: in this case, it has been shown that the interplay between structure and dynamics leads to some universal behavior that can be exploited to capture the most fundamental mechanisms of information exchange among interconnected units [79]. To correctly understand communication and correlation in networks, therefore, it is crucial to understand that not only shortest path but also many more routes—not necessarily the optimal ones—contribute to information pathways: this fact is well captured, for instance, by network communicability [80–82] (see [83] for a thorough review), which has been recently used to determine the informational cost of navigating a network under different levels of external noise and, as an application,
to determine the levels of noise at which a protein–protein interaction (PPI) network seems to work in normal conditions in a cell [84].

In the remainder of this work we will focus on a recently proposed approach which explicitly accounts for information dynamics and provides a solution to the sub-additivity problem, by using a quantum-like density matrix, defined as a Gibbs state [85]. Despite a physical interpretation of this density matrix in terms of the propagator of diffusion-like dynamics has been provided, more recently it has been shown that it corresponds to a special case of a more general statistical field theory of complex information dynamics [86], providing a potentially unifying solution to the two major challenges outlined at the beginning of this section, which will be described in more detail in the next section.

3. Field theory of complex information dynamics

In a field theory, the physical quantity of interest can be measured at each point in space and time: in the case of networks, space is encoded by system’s units, mapped to a finite number of points in the Euclidean space (see [32] for details). In the following, the physical quantity of interest is generally regarded as the stream of information (e.g., water, electrochemical signals, human flows, so forth and so on) which is an inherently fluctuating object, making approaches based on classical field theory not suitable to describe its state and its dynamics. Instead, it is

\[ W_{i,j} \]

where the amount of field on top of node \( i \) at time \( \tau \) reads \( \langle i|\phi(\tau)\rangle \). The evolution of the information field in the most general form is given by

\[
\partial_\tau |\phi(\tau)\rangle = \mathbf{F}(\tau, \mathbf{W}, |\phi(\tau)\rangle),
\]

which after linearization becomes

\[
\partial_\tau |\phi(\tau)\rangle = -\mathbf{H}|\phi(\tau)\rangle,
\]

with \( \mathbf{H} \) playing the role of a control operator. Equation (2) directly leads to

\[
|\phi(\tau)\rangle = \mathbf{S}(\tau)|\phi(0)\rangle,
\]

with the propagator \( \mathbf{S}(\tau) = e^{-\tau \mathbf{H}} \). \( \mathbf{H} \) can describe a range of possible dynamics of the information field, such as continuous-time diffusion, random walks, consensus and synchronization on top of single layer and multilayer networks.

Taking \( \phi_0 \) to be the initial value of the field on top of node \( i \), the information flow from the node can be obtained as

\[
|\phi(\tau)\rangle = \phi_0 S(\tau)|i\rangle,
\]

and the information received by \( j \)th node from \( i \)th follows \( \phi_0 \langle j|S(\tau)|i\rangle \). To provide a statistical description of the information dynamics, we assume a probabilistic initial condition

\[
|\phi(0)\rangle = \sum_{i=1}^{N} p_i |\phi_0 i\rangle,
\]

where \( p_i \) is the probability that \( i \)th node is the information field seed. Accordingly, the expected information flow from \( i \)th node becomes \( p_i \phi_0 S(\tau)|i\rangle \) and the expected information received by \( j \)th node from \( i \)th node follows \( p_i \phi_0 \langle j|S(\tau)|i\rangle \).

In absence of information about the initiator node, we assign uniform probabilities \( p_i = 1/N \), leading to the expected information flow from \( i \)th node \( \frac{\phi_0}{N} S(\tau)|i\rangle \), and to the expected information received by \( j \)th node from the \( i \)th one given by

\[
\frac{\phi_0}{N} \langle j|S(\tau)|i\rangle.
\]
traps is equal to the size of that stream \( \hat{\phi}_n \) and, consequently, the overall expected trapped field can be obtained from the summation of stream sizes \( \sum_{n=0}^{N} \hat{\phi}_n \ell_n(\tau) = \hat{\phi}_N Z(\tau) \), where \( Z(\tau) = \sum_{n=1}^{N} \ell_n(\tau) = \text{Tr}[\hat{S}(\tau)] \).

Since the expected trapped field regulates the size of information streams, it can be considered responsible for activation of the streams. Therefore, the information dynamics is reducible to the dynamics of the trapped field. Consequently, using a proper superposition of information streams, \( \hat{\rho}(\tau) = \sum_{n=1}^{N} \ell_n(\tau) \mathcal{S}^{(n)} = \frac{\hat{\phi}_N}{N} \hat{Z}(\tau) \), where \( \ell_n(\tau) \) is the \( n \)th information stream and \( \mathcal{S}^{(n)} \) is the corresponding potential density matrix for the system. Assume the field is discretized into a large number of infinitesimal quanta carrying value \( h \) which, depending on the nature of information field, can be bits of information, small packets of energy, infinitesimal volumes of fluid, etc. A number \( n(\tau) = \frac{\hat{\phi}_0}{N} \hat{Z}(\tau) \) of quanta activates the information streams and, similarly, the number \( n^{(\ell)}(\tau) \) of quanta participating in activation of \( \ell \)-th information stream is given by \( n^{(\ell)}(\tau) = \frac{\hat{\phi}_0}{N} \ell_n(\tau) \). Thus, the probability that one quantum participates in activation of \( \ell \)-th stream is given by \( \frac{n^{(\ell)}(\tau)}{n(\tau)} = \frac{\hat{\phi}_0}{N} \ell_n(\tau) = \ell(\tau) \) and each quantum of the trapped field generates a unit flow

\[
\frac{1}{n(\tau)} \frac{\hat{\phi}_0}{N} \hat{Z}(\tau) \hat{\rho}(\tau) |\bar{i}\rangle = h \hat{\rho}(\tau) |\bar{i}\rangle,
\]

by activating one of the information streams according to their activation probabilities. The unit flow is the smallest element used to describe functional interactions in the system. The expected information flow is obtained from summation of all the unit flows.

The information dynamics can be fully captured using a number of unit flows, each activating one of the information streams \( \{\mathcal{S}^{(n)}(\tau)\} \) (see figure 2). Because of their probabilistic nature, the streams shape a statistical ensemble encoding all possible fluxes among components and their probabilities and the operator \( \hat{\rho}(\tau) \) is reminiscent of the density matrix used in quantum statistical physics in terms of the superposition of quantum states, with the important distinction that here we do not necessarily deal with physical quantum objects.

Consequently, \( Z(\tau) \) that determines the amount of field trapped on top of the initiator units can be interpreted as partition function. When the propagator is derived for random walk dynamics \( (H = \hat{L}^\dagger) \), \( Z(\tau) = NR(\tau) \), where \( R(\tau) \) is the average return probability of random walkers. Average return probability is a proxy for the tendency of structure to trap the flow in its initial place. Therefore, the partition function \( Z(\tau) \) can be interpreted as dynamical trapping and used to assess the transport properties of networks. Of course, the redundancy and symmetries in diffusion pathways for information propagation can have negative effect on signal transport (see figure 3)—e.g., a regular lattice provides slow information propagation, while

![Diagram of information field components](image-url)

**Figure 2.** Information field components. Information streams for a network of three connected nodes, where the dynamical process is classical random walk. It is shown that the activation probabilities of streams change over time. Blue and red arrows, respectively, positive and negative fluxes between the nodes, whereas, self-loops trap the field on top of the initiator. Reprinted figure with permission from [86], Copyright (2020) by the American Physical Society.
long range interactions between distant nodes and topological complexity can lower the dynamical trapping and enhance the transport in small-world or scale-free networks (see figure 3).

Moreover, in a wide range of complex systems the structure is better represented in terms of multilayer networks [38], where the interactions between the units is described by $L > 1$ distinct networks, seen as layers, coupled together. The existence of multiple interaction types in such systems is important for a range of properties. Using the statistical physics of complex information dynamics, one can obtain a fundamental inequality that relates the dynamical trapping of the multilayer system and the geometric average of partition functions of layers, through average dynamical trapping:

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

where $Z(\tau)$ is the partition function of the multilayer system and $Z^{(\ell)}(\tau)$ is the partition function of the $\ell$th layer. The inequality is important as it proves the advantage of multilayer systems over single layers, according to information transport. It is worth mentioning that the concept of dynamical trapping has been generalized to other types of dynamical processes [87].

4. Network information-theoretical tools and distances

Information-theoretic measures have been successfully applied in a wide variety of disciplines, from biology to quantum physics. The aim of this section is to review such measures developed to study complex networks, density matrices obtained from physical processes on top of networks used to quantify their information content, as well as the (dis)similarity metrics exploited to compare structured data. The applications of the mathematical objects presented here to real-world complex networks will be discussed in the following sections.
4.1. Spectral entropies, divergences and likelihood

The probability distribution $\rho_\ell(\tau)$ is crucial for the information flow between the nodes. The mixedness of information streams can be quantified by the von Neumann entropy of the density matrix:

$$S(\tau) = -\operatorname{Tr}[\hat{\rho}(\tau) \log \hat{\rho}(\tau)],$$

as information dynamics becomes rich and diverse, the number of information streams required to capture it grows, and consequently, entropy takes higher values, even at large propagation times. It is worth noting that when the continuous diffusion governs the flow, the von Neumann entropy of the ensemble coincides with the spectral entropy introduced to quantify the information content of networks [85].

It is worth noting that a more general definition of network entropy is the Rényi entropy

$$S_q(\tau) = \frac{1}{1-q} \log \operatorname{Tr}[\hat{\rho}^q(\tau)],$$

which becomes the Hartley entropy in the case $q = 0$, recovers the von Neumann entropy (equation (7)) as $q \to 1$ and equals the collision entropy when $q = 2$ [88].

Remarkably, the von Neumann entropy can be used to quantify the (dis)similarity of networks, from the perspective of information theory. Assume $\hat{\rho}(\tau)$ is the density matrix of a network $G$ and $\hat{\rho}'(\tau)$ is the density matrix of another network $G'$, at specific temporal scale $\tau$. The Kullback–Leibler entropy divergence between the two is given by

$$D_{KL}(\hat{\rho}(\tau)\|\hat{\rho}'(\tau)) = \operatorname{Tr}[\hat{\rho}(\tau)(\log \hat{\rho}(\tau) - \log \hat{\rho}'(\tau))]$$

which quantifies the amount of information gained about $G$, by observing $G'$. For instance, this is useful when the density matrix of network $G'$ is used as a model to predict $\hat{\rho}(\tau)$ and the entropy divergence quantifies the prediction error. The Kullback–Leibler entropy divergence is non-negative ($D_{KL}(\hat{\rho}(\tau)\|\hat{\rho}'(\tau)) \geq 0$), and not symmetric ($D_{KL}(\hat{\rho}(\tau)\|\hat{\rho}'(\tau)) \neq D_{KL}(\hat{\rho}'(\tau)\|\hat{\rho}(\tau))$). Interestingly, by minimizing the Kullback–Leibler entropy divergence of real-world networks and parametric network models, one can infer the optimal parameters describing the observed networks through maximum-likelihood estimation (see figure 4), noting that maximizing likelihood is equivalent to minimizing the Kullback–Leibler divergence [89] (for a proof in case of network density matrices see reference [85]). The likelihood is given by the $\mathcal{L}(\hat{\rho}(\tau)) = \|\hat{\rho}(\Phi, \tau)\|$, where $\Phi$ indicates the parameter(s) to be optimized, $\hat{\rho}(\Phi, \tau)$ is the density matrix of network model given the parameters and $\hat{\rho}(\tau)$ is the density matrix of an empirical network. Often, the suitable model for real data is obtained in a trade-off between maximizing a function, which in this case is the likelihood, and minimizing the number of parameters used to construct the model. Akaike information criterion, given by $2g - 2 \log_{\gamma} \mathcal{L}$ where $g$ is the number of parameters and $\mathcal{L}$ is the likelihood function, can be used for such optimization [85].

The Jensen–Shannon divergence is given by

$$\mathcal{J}(\hat{\rho}(\tau)\|\hat{\rho}'(\tau)) = \frac{D_{KL}(\hat{\rho}(\tau)\|\hat{\rho}'(\tau)) + D_{KL}(\hat{\rho}'(\tau)\|\hat{\rho}(\tau))}{2},$$

whose square root provides a symmetric metric for quantifying network dissimilarity, at variance with the Kullback–Leibler divergence. In the analysis of multilayer networks, the Jensen–Shannon distance has been used to quantify the pairwise dissimilarity of layers (see figure 4) unravelling their hierarchical organization [85].

4.2. Information-theoretic distances and dimensionality reduction

As mentioned previously, in many scenarios, the underlying structure of a complex system can be better represented as a multilayer network, where each layer is a network describing a particular type of interaction between the nodes. In these cases, a quantity of interest is the average Kullback–Leibler entropy divergence between the multilayer system and its layers, that is named intertwining and measures the difference between the whole and its parts (layers), from the perspective of information theory [90]. Remarkably, intertwining is proxy of the overall redundancy (or similarity) of layers, having low values when the layers are highly similar and high values for diverse layers. Thus, if there is a subset of layers highly similar to each other, their redundancy can be identified and quantified. In a multilayer network with $L$ layers, quantifying the overall diversity of layers using intertwining and identifying groups of similar layers [91], using Jensen–Shannon divergence allows for devising dimensionality reduction algorithms aiming to merge the redundant layers and return a multilayer network with $L' < L$ maximally diverse layers (see figure 8). In the following sections, we discuss an application of dimensionality reduction to empirical systems.
Figure 4. Distances, fitting and hierarchical clustering. (A) Heatmap of the Jensen–Shannon distance between layers of the human microbiome, with $\tau = 0.1$. The hierarchical clustering is shown on top of the heatmap. (B) An example of fitting parameters using Kullback–Leibler minimization and maximum likelihood approach. More specifically, a Watts–Strogatz’s model ($N = 200$) for different average degree $K$ and rewiring probability $P_{\text{rew}}$ is compared against a Watts–Strogatz network with $K = 6$ and $P_{\text{rew}} = 0.2$, assumed to be the data to fit. Reproduced from [85]. CC BY 4.0.

5. Applications to empirical interconnected and coupled systems

Being grounded in information theory and statistical physics, the presented framework has broad applications to empirical systems across scales, from within cells to complex societies. In the following, we briefly review a few applications, from clustering analysis of human-viral interactomes and human microbiome, to the centrality, robustness analysis, the dimensionality reduction and transport phenomena in social and transportation systems.

5.1. Multiscale analysis of information dynamics

The dependence of the density matrix $\hat{\rho}(\tau)$ on the propagation time $\tau$ of the dynamics, allows one to use $\tau$ as a resolution parameter. At the lowest values ($\tau \approx 0$) the microscale is explored, at the largest values ($\tau \geq N$, where $N$ is the system size) the macroscale is analyzed, whereas at the intermediate values the mesoscale is probed. Alternatively, since the temporal scale can be different from network to network, depending on the connectivity, number of nodes, topology, etc, one can use the diffusion time to allow for comparison across networks. Let the eigenvalues of the control operator $\hat{H}$ be $\lambda_\ell (\ell = 1, 2, \ldots N)$, ordered as $\lambda_1 \leq \lambda_{\ell+1}$ and $\lambda_1 = 0$. Naturally, the eigenvalues of the propagator $\hat{S}(\tau)$ follow $e^{-\tau \lambda_\ell}$, ($\ell = 1, 2, \ldots N$). The second eigenvalue of the control operator determines the diffusion time $\tau_d = 1/\lambda_2$—i.e., the temporal scale close to equilibrium. One can divide the propagation time scale $\tau$ by the diffusion time to obtain the rescaled temporal parameter $\tau/\tau_d$ allowing for comparisons across network types and sizes [92]. Note that the control operator can take the shape of operators other than combinatorial Laplacian and, in those cases, one can use the definition of diffusion time as long as the control operator has non-negative eigenvalues, and at least one eigenvalue that is zero. To date, two distinct classes of applications have been introduced for the analysis of complex information dynamics across scales:

- **Density state** (or Gibbs state): the network state is given by $\hat{\rho}(\tau)$ and descriptors are obtained from calculating the corresponding von Neumann and relative entropies for varying $\tau$;
- **Emergent functional state**: the network state evolves, allowing one to analyze each state at time $\tau$ as the temporal snapshot of a time-varying process. Accordingly, at each time $\tau$ a functional network emerges that encodes the pairwise flow between the nodes, and can be analyzed by means of classical network descriptors (see figure 5).

**Analysis of density states.** Often, components of complex systems might appear structurally similar while they take distinct functional roles. For instance, cells active in the visual cortex and auditory cortex of human
Figure 5. Multiscale analysis. (A) The emergent functional state corresponding to a *Physarum polycephalum* (Pp) fungal network, at five different scales $\tau = 0.001, 0.009, 0.079, 0.708, 6.31$, where each functional module, captured by the Louvain algorithm, is colored differently. Reproduced with permission from [92]. (B) Von Neumann entropy as a function of $\beta$ (here $\beta$ indicates the temporal scale $\tau$) for a highly-ordered network colored in orange, and its configuration model (CM) colored in blue. At small temporal scales, the von Neumann entropy reaches its maximum and in the large limit of the propagation time, it is related to the number of connected components $C$. At the mesoscales, the entropy is able to capture the mesoscopic organization and the height of the plateau is related to the overall modularity of the network. Reproduced from [93]. CC BY 4.0.

brain are neurons, yet they are involved in different computational tasks. The reason behind such a diverse functionality of similar agents might be their position with respect to the structure, as it partly determines how these agents handle information, as senders, receivers and processors. Thus, one way to quantify the functional diversity of a system is to identify the functional modules—i.e., groups of nodes that exchange information mostly within themselves—that will be discussed in the following. Alternatively, one can quantify the diversity of flow distribution vectors originated from the nodes, in terms of the corresponding average cosine distance. It has been shown that such measures of functional diversity are proportional to the von Neumann entropy of the system, in synthetic and empirical networks [86, 92, 94].

Using this framework, one can analyse the virus-human PPI as an interdependent system with two parts, human PPI targeted by viral proteins, and can quantify the effect of each viral infection on the information dynamics, dynamical trapping and the functional diversity of units on the human interactome, and map the perturbations caused by a virus to compare the effects of distinct viruses gaining insights about the systemic effects of SARS-CoV-2 [95].

The framework has been also used to study the effect of connectomes’ topological complexity on information dynamics within the human brain. More specifically, real connectomes from healthy subjects have been compared against randomized network models, characterized by distinct structural features, such as the Erdős–Rényi model (ERM), the CM, the stochastic block model (SBM) and the hyperbolic model (HM). Comparing these generative models with the real data sheds light on the properties of human connectomes. For this comparison, classical and maximal entropy random walks (MERW) have been used to construct the density matrices: interestingly, most generative models have smaller von Neumann entropy than empirical human brains between meso and macroscopic temporal scales, where mid- or long-range communications between brain areas take place (see figure 6).
A similar investigation has been done on three species of fungi and slime molds, including Pp, Phanerochaete velutina (Pv) and Resinicium bicolor (Rb). The networks corresponding to these species are weighted, according to the cord conductance for pairwise interactions (see [96] for details about the data set). Similarly, to connectomes, the von Neumann entropy of these biological networks is larger than their randomized models at the meso-scale (see figure 6).

Analysis of emergent functional states. In contrast with the adjacency operator, that gives a static and localized picture of interactions between the nodes, the propagator encodes pairwise flow exchange between them at multiple scales, characterized by a temporal parameter $\tau$ describing short-, middle- and long-range functional interactions. Thus, one can define the emergent functional state of the system as $\hat{\omega}(\tau) = f[\hat{\rho}(\tau)] = Z(\tau)\hat{\rho}(\tau) \circ (U - I)$, where $I$ is the identity matrix, $U$ is the matrix with all entries equal to 1, and $\circ$ denotes the Hadamard product. Operatively, the operator $\hat{\omega}(\tau)$ is equivalent to the propagator of information dynamics with diagonal entries set equal to 0 and $\langle j|\hat{\omega}(\tau)|i \rangle$ gives the flow received by node $j$ from node $i$ at time $\tau$. Interestingly, it can be shown that for continuous diffusion ($\hat{H} = \hat{L}$) at extremely small $\tau$, the emergent functional state reduces to the adjacency operator.

Emergent functional states $\hat{\omega}(\tau)$ can be analyzed to identify functional modules—i.e., groups of nodes that exchange higher amount of flow between themselves than with the rest of the nodes—using the Louvain algorithm. In the case of the fungal networks described above, it has been shown that for short-range communications are considered, a large number of functional modules are detected, and for long range interactions a smaller number of modules emerge (see figure 5). The number of modules is proportional to the von Neumann entropy of the system [92].

5.2. Network entanglement as a proxy for robustness

As discussed in the previous sections, the von Neumann entropy measures the diversity of information dynamics in the system and functional diversity of its units. It is possible to quantify the importance of any node $x$ for the functional diversity of the system, in terms of the effect of node removal on the von Neumann entropy. The network before the detachment of node $x$ can be denoted as $G$, the remaining part of network after detachment is shown as $G'_x$ and the star network containing the detached node and its links $\delta G_x$. Here, we denote their von Neumann entropies, respectively, as $S(\tau)$, $S'_x(\tau)$ and $S_x(\tau)$. The definition of entanglement [87] between node
Figure 7. Node network entanglement. (1) The detachment of node \( x \) and its emanating edges from the original network \( G \) results in a perturbed network \( G' \) and a star network \( \delta G_x \) colored in red. (2) In a random geometric network \((N = 100 \text{ and radius } 0.15)\), the entanglement corresponding to each node is shown as a function of the propagation time \( \beta \) where each trajectory is colored according to the degree of the detached node. The average entanglement \( \bar{M}_\beta \) is represented by the orange dashed line. (3-a) Disintegration of different network types is considered. The robustness of an ensemble of each network model is tested against random failures and targeted attacks based on a variety of centrality measure including entanglement at three temporal scales (-small, -mid, -large), is shown. (3-b) Entanglement centrality, tuned at relatively large propagation time-scale, performs equal or faster than other measures in breaking the network up to its critical fraction (b). Reproduced from [87]. CC BY 4.0.

For \( x \) and the network is given by

\[
M_x(\tau) = [S'_x(\tau) + S_x(\tau)] - S(\tau).
\]

(12)

It is possible to provide an analytical understanding about the behavior of entanglement, by using a mean-field approximation of the von Neumann entropy, in the case of continuous diffusion process \((\hat{H} = \hat{L})\), as

\[
S_{\text{MF}}(\tau) = \tau \bar{k} \frac{Z(\tau) - 1}{Z(\tau)} + \log Z(\tau),
\]

(13)

where \( \bar{k} \) is the average degree. Using the mean-field entropy, one can show that at extremely small and large temporal scales, the entanglement follows:

- \( \tau \to 0: M_x(\tau) \approx \log(k_x + 1) \)
- \( \tau \to \infty: M_x(\tau) \approx \tau \bar{k} \log C'_x \)

Where \( k_x \) is the degree of the removed node and \( C'_x \) is the number of disconnected components, in the perturbed network \( G'_x \). At the meso-scale, entanglement assesses the importance of nodes for the transport properties of the system, by measuring its effect on the dynamical trapping. As an application, entanglement has been used as a centrality measure capturing the role played by nodes in keeping the overall diversity of the information flow. It has been shown that attack strategies based on this centrality measure at the meso-scale, are compatible with or outperforms other methods in driving empirical social, biological and transportation systems to fast disintegration, showing that the nodes central for information dynamics are also responsible for keeping the network integrated [87] (For more details see figure 7).
5.3. Reducing complexity without altering the structure

As mentioned previously, information-theoretic metrics can be used to quantify the distance between pairs of networks. Interestingly, the Jensen–Shannon distance (equation (10)) has been used to assess the dissimilarity of layers of the multilayer network corresponding to the sites of the human microbiome, leading to their hierarchical clustering, in agreement with the state-of-the-art community-based association methods [85].

While the multilayer representation often provides a more accurate framework to model the structure of complex systems, it is a challenge to find the minimum number of layers necessary to precisely represent the structure. One way to tackle the problem is to find the layers containing redundant information, merge them together and maximize the distinguishability between the multilayer and the aggregated graph, obtained from merging all the layers simultaneously. To identify the redundant layers, one can calculate their hierarchical clustering given by means of the Jensen–Shannon distance and devise a procedure to merge them until the layers are maximally diverse and distinguishable [91]. However, the proposed reducibility algorithm had shortcomings. For instance, in case of synthetic multilayer systems where a subset of layers are identical, and maximally redundant, the framework does not lead to the aggregation of those layers. Moreover, the formalism cannot be used to understand the effects of dimensionality reduction on the dynamical properties of the system.

To overcome those issues, instead of relying on the layer distinguishability, one can maximize the intertwining [90]: the average entropy divergence between the multilayer and its layers. This function has an elegant interpretation from the perspective of complexity science, as it directly quantifies the importance of being a multilayer system, in terms of the difference between the whole (multilayer) and its parts (layers). It has been shown that by maximizing intertwining, one can effectively reduce the redundancies, reaching maximally diverse layers, enhancing the transport properties of systems, lowering the dynamical trapping and increasing the navigability. Interestingly, to apply this framework to real-world multilayer networks, it is not necessary to alter their structure: it is enough to couple the dynamical processes on top of similar layers to boost the transport in the systems. For instance, instead of adding new routes, one can introduce shared bus and subway tickets to couple the layers and enhance transport properties. For more details on the enhancement of transport properties in empirical systems including co-authorship network of scientists, European airport networks and the London public transportation network as a result of the functional dimensionality reduction, see figure 8.
6. Future perspective

Recently, a promising theoretical framework has been developed to reconcile thermodynamics and information to quantify, for instance, the energetic cost of altering a system’s state. This novel field, based on stochastic thermodynamics and fluctuation theorems, is increasingly gaining attention for its applicability to a broad spectrum of problems which cannot be easily tackled with existing frameworks [97]. Even more recently, a time-information uncertainty relation bounding the rates of energy and entropy exchange has been discovered [98], the thermodynamics of modular computations has been linked to structural energy cost [99], the thermodynamic cost of Turing machines—emblematic computing models—quantified [100], and experimental evidence linking information with thermodynamics has been provided [101], supporting the fact that information is physical, as originally conjectured by Landauer et al [102].

A traditional way to deal with information relies on Shannon’s theory of communication [103], building on the concepts of data source, communication channel (through which the data are transmitted, being possibly contaminated by noise) and data receiver. In this setup, the information content is often quantified in terms of entropy, encoding the ability of the receiver to reconstruct from the observed signal the original data sent by the source. For this reason, Shannon’s entropy is, nowadays, one of the widest used measure to characterize the regularity of patterns and, in some cases, as a proxy for their complexity. We have mentioned some representative studies based on information entropy and its variants (e.g., relative entropies) to characterize network complexity, although their estimation—limited by the sub-set of network descriptors they focus on—does not fully exploit the richness of the structural data. It is worth mentioning that the relation between a graph complexity and its information is a problem older than network science itself, known as structural information content, which has been already faced a few years after Shannon’s pioneering work (e.g., see [104–107]). Since the last decade, information theoretic approaches are being developed in network science, using source-channel-receiver paradigm [108], where analyses such as coarse-graining and identification of communities are reliably mapped into the classical problem of decoding a message transmitted along a noisy channel [109–112].

In data analysis, dealing with information is crucial for several practical reasons. For instance, an adequate quantification of network information content allows one to compare systems of different types (e.g., a biological against a technological one) and of different sizes, a task that nowadays can be performed with promising approaches—from quantum Jensen–Shannon divergence [85] to network portrait divergence [113], to mention a few ones [114]—under some limitations peculiar of each method. Another practical application would concern the quantification of how a node influences and is influenced by the network [115], as well as to gain insights on how information content is related to system’s robustness [116].

The research agenda for the next future is rich. While, on the one hand, the framework presented in the previous chapters has proven to be effective for practical applications like network comparison, a current technical limitation, on the other hand, is that only graphs of the same size can be compared: the possibility to extend the formalism to network states of different size will favor applications to detect relevant changes of information in time-varying systems, among others. The lack of consensus on how to measure network complexity and approaches to discriminate between different networks in terms of their structure, information capacity and processing, is another promising direction. For instance, existing approaches to measure network efficiency in information exchange [117, 118] might be better understood within the statistical field theory of information dynamics. While a generalization of the presented framework to the case of multilayer systems has been recently introduced [90], its extension to the realm of higher-order models [42, 43] of complex systems is still missing. It is worth mentioning that other theoretical developments and applications recently appeared in the literature. For instance, Nicolini et al exploited the analogy with thermodynamics to provide a physical interpretation to inference applied to empirical brain networks [119] and to assess the effects of noise and image processing on functional connectivity [93]. Su et al have proven that spectral entropy is able to identify variations in network topology from a global perspective better than traditional distribution entropy [120], while Glos et al have shown that the phase transition in the resolution parameter of the spectral entropy of empirical systems can be used to distill the information whether the graph represents real-world interactions [121].

Furthermore, it has been shown that there is a strict relationship between information entropy, network robustness and network curvature in terms of the so-called Ollivier–Ricci curvature [122]. Since robustness can be interpreted as the rate function at which a network returns to its original state after a perturbation, it is positively correlated with entropy and, through entropy, it is related to graph curvature [123], providing an exciting opportunity for our framework to be extended to characterize system’s robustness and its graph curvature. The possibility to define a suitable network-based Fisher–Rao metric might allow one to exploit the geometry of statistical manifolds to characterize system complexity and to provide a ground for machine learning applications based on information geometry [124], which has been successfully related to critical and phase transition phenomena in classical statistical mechanics [16, 125–127] (see [128] for a review). In the
long term, the framework has the potential to provide a unifying ground for an information theory of complex networks, with opportunities for novel machine learning techniques, and for the analysis of information in quantum networks. An even more exciting perspective concerns with its cross-pollination with other theoretical frameworks (see figure 9), such as stochastic thermodynamics, with the aim to gain insights about how real-world complex systems handle information and its interplay with energetic cost.

7. Conclusions and discussion

In this work we have briefly described the importance of network science for modeling and analyzing empirical interconnected systems, as well as its current role in apparently disconnected research fields, such as statistical physics, quantum information and technologies, and machine learning. In fact, network science can be equally well characterized by its theoretical contributions to complexity and systems science as well as by the algorithmic and computational tools it provided to gain insights from structured data. This dual nature of network science has been, and it is still, beneficial for its development and its cross-pollination with other fields, from biology to social sciences, as well as for applications.

In the journey towards the development, on the one hand, of theories able to reproduce the most salient features of complex systems, such as the emergence of collective behavior and critical phenomena, and, on the other hand, of analytical tools able to fully exploit the richness of structural data, the concept of information is increasingly gaining momentum. In fact, a complex network can be thought in terms of units which process and exchange information between each other, driving the system from an input state to an output state. With respect to this point of view, complex systems might resemble one another in the way they handle information, as also conjectured by Gell–Mann. Therefore, including information within a grounded and systematic theoretical framework allowing for direct applications to empirical data is of paramount importance and one of the most challenging tasks of the next future. For instance, it has been shown that the search strategy of macroscopic searchers with sparse information can be reliably described by ‘infotaxis’, a process which locally maximizes the expected rate of information gain [129]. Interacting living systems can dynamically adapt to be more efficient in dealing with heterogeneous and uncertain environmental conditions when they operate at criticality, which in turn can emerge from information-based fitness [130]. Biochemical systems generate a highly diverse set of complex molecules which exploit processes able to decode and encode information which constrains and drives reaction networks [131].
Here, we have reviewed in some detail the theoretical background and the applications of a framework recently developed to overcome some of the limitations described above, which is grounded on a statistical field theory of complex information dynamics. The main theoretical object of this framework is the network density matrix, encoding the ensemble of information operators—known as streams—which are responsible for information flows within the network. At variance with other matrix functions, the density matrix described in this work allows to (i) describe a network state as the interplay between the underlying topology and the information dynamics on the top of it, and (ii) define an information entropy, as a suitable classical counterpart of the quantum von Neumann entropy widely used in quantum thermodynamics and quantum information science. This framework allows for the definition of a variety of information-theoretic tools such as relative entropies or divergences, as well as information-theoretic distances, e.g., the Jensen–Shannon divergence—which can be reliably used to compare networks of the same size, such as the layers of a multiplex system [32]. Being grounded on density operators and von Neumann entropy, we expect this framework to have the same potential it has for quantum mechanical information theory, where it already provided a unified description of classical correlation and quantum entanglement [132]. To highlight its potential, we have described some applications of this framework to characterize system’s robustness to random and targeted disruptions, as well as to perform dimensionality reduction of coupled systems, with applications ranging from the human interactome to the human brain at different stages of dementia, from a web of scientific collaborations to the backbone of the London tube.

Finally, this work presents possible future research perspectives to expand the framework of complex information dynamics, from finding more general distance measures for network comparison and defining operators quantifying the topological complexity, to exploring more bridges that might connect the framework with realms of knowledge such as non-equilibrium thermodynamics and machine-learning.

Acknowledgments

The authors thank Barbara Benigni for kindly providing a visualization of the human brain network used in this work, Gemma de las Cuevas and Artemy Kolchinsky for interesting discussions.

Data availability statement

No new data were created or analysed in this study.

ORCID iDs

A Ghavasieh https://orcid.org/0000-0001-8138-7208
M De Domenico https://orcid.org/0000-0001-5158-8594

References

[1] Watts D J and Strogatz S H 1998 Nature 393 440–2
[2] Barabási A L and Albert R 1999 Science 286 509–12
[3] Guimerà R and Nunes Amaral L A 2005 Nature 433 895–900
[4] Bullmore E and Sporns O 2012 Nat. Rev. Neurosci. 13 336–49
[5] Allard A and Serrano M A 2020 PLoS Comput. Biol. 16 e1007584
[6] Watts D J, Dodds P S and Newman M E J 2002 Science 296 1302–5
[7] Dodds P S, Muhamad R and Watts D J 2003 Science 301 827–9
[8] Boguñá M, Papadopoulos F and Krioukov D 2010 Nat. Commun. 1 62
[9] Barabási A-L, Gulbahce N and Loscalzo J 2011 Nat. Rev. Genet. 12 36–68
[10] Zhou X, Menche J, Barabási A L and Sharma A 2014 Nat. Commun. 5 4212
[11] Mencic J, Sharma A, Kitsak M, Ghissian S D, Vidal M, Loscalzo J and Barabási A L 2015 Science 347 1257601
[12] Verbavatz V and Barthélémy M 2020 Nature 587 397–401
[13] Montoya J M, Pimm S L and Solé R V 2006 Nature 442 259–64
[14] Pocock M J O, Evans D M and Memmott J 2012 Science 335 973–7
[15] Suweis S, Grilli J, Banavar J R, Allesina S and Maritan A 2015 Nat. Commun. 6 10179
[16] Grilli J, Adoríos M, Suweis S, Barabási G, Banavar J R, Allesina S and Maritan A 2017 Nat. Commun. 8 14389
[17] Ballerini M et al 2008 Proc. Natl Acad. Sci. 105 1322–7
[18] Wild B, Dormagen D M, Zachariae A, Smith M L, Traynor K S, Brockmann D, Coulizzi D and Landgraf T 2021 Nat. Commun. 12 1110
[19] Holme P and Saramäki J 2012 Phys. Rep. 519 97–125
[20] Kleineberg K-K, Boguñá M, Ángeles Serrano M and Papadopoulos F 2016 Nat. Phys. 12 1076–81
[21] Bianconi G and Rahmede C 2017 Sci. Rep. 7 41974
