Algorithmic Complexity of Multiplex Networks

Andrea Santoro[1,2] and Vincenzo Nicosia[1,*]

[1]*School of Mathematical Sciences, Queen Mary University of London, London E1 4NS, United Kingdom*
[2]*The Alan Turing Institute, The British Library, NW1 2DB, London, United Kingdom*

Multilayer networks preserve full information about the different interactions among the constituents of a complex system, and have recently proven quite useful in modeling transportation networks, social circles, and the human brain. A fundamental and still open problem is to assess if and when the multilayer representation of a system provides a qualitatively better model than the classical single-layer aggregated network. Here we tackle this problem from an algorithmic information theory perspective. We propose an intuitive way to encode a multilayer network into a bit string, and we define the complexity of a multilayer network as the ratio of the Kolmogorov complexity of the bit strings associated to the multilayer and to the corresponding aggregated graph. We find that there exists a maximum amount of additional information that a multilayer model can encode with respect to the equivalent single-layer graph. We show how our complexity measure can be used to obtain low-dimensional representations of multidimensional systems, to cluster multilayer networks into a small set of meaningful superfamilies, and to detect tipping points in the evolution of different time-varying multilayer graphs. Interestingly, the low-dimensional multiplex networks obtained with the proposed method also retain most of the dynamical properties of the original systems, as demonstrated for instance by the preservation of the epidemic threshold in the multiplex susceptible-infected-susceptible model. These results suggest that information-theoretic approaches can be effectively employed for a more systematic analysis of static and time-varying multidimensional complex systems.

I. INTRODUCTION

The success of network science in modeling real-world complex systems [1,2] relies on the hypothesis that the interconnections among the elementary units of a system—i.e., the network of their interactions—are responsible for the emergence of complex dynamical behaviors [3,4]. Traditionally, relevant contributions toward a better understanding of complex networks have come from statistical physics [5–7], where the main aim is to characterize the ensembles of random graphs comparable with an observed real-world network. However, really interesting results have also come from information theory. A quite prolific line of research in this area aims at adapting classical concepts and methods from information theory to networks analysis [8–10]. Some other studies have focused instead on the definition of entropy measures on empirical networks [11,12] and on the quantification of the significance of structural indicators based on algorithmic information theory [13,14].

Multilayer and multiplex networks, which take into account different kinds of relations among the same set of nodes at the same time [15–17], are a currently hot research topic in network science. The main idea behind the investigation of high-dimensional network representations is that retaining full information about the structure of a system under study is often fundamental to fully understand its behavior. Indeed, multilayer networks have helped unraveling interesting structural properties in transportation systems [18,19] and neuroscience [20,21], and have revealed qualitatively new emerging phenomena, including abrupt cascading failures [22], superdiffusion [23], explosive synchronization [24], and hyperfast spreading [25–28].

These encouraging results have transformed our understanding of many physical systems, but an overarching question remains about whether it is necessary to incorporate all the available data about a system in order to fully characterize its behavior [29]. Some recent studies have indeed shown that the multilayer version of some dynamical processes cannot be reduced to the corresponding single-layer process on any simple combination of the existing layers [30]. Nevertheless, determining whether a lower-dimensional multilayer network can exhibit the same

*[*]v.nicosia@qmul.ac.uk

structural and dynamical richness of the full multilayer graph is still an open question. Some concrete attempts to solve this problem have come from a formalization of multilayer dimensionality reduction in terms of a quantum information problem [31] and from other approaches relying on mesoscopic similarity between layers [32–35]. However, we still lack a convincing method to quantify the amount of information contained in a multilayer network model and to compare the information content of different multilayer networks.

In this paper, we take an algorithmic complexity perspective on this problem, and we define the complexity measure $\mathcal{C}(\mathcal{M})$ to quantify the amount of information contained in a multiplex network $\mathcal{M}$. The measure leverages the classical concept of Kolmogorov complexity [36], according to which the complexity of a bit string is equal to the length of the shortest possible program that can produce that string as its output. In particular, we show that $\mathcal{C}(\mathcal{M})$ is quite useful in determining the optimal number of layers needed to represent a multilayer network and in detecting structural and dynamical similarities among multilayer networks from different domains.

II. RESULTS

We propose here a formalism to quantity the complexity of a multiplex network over N nodes and M layers, based on the comparison of the Kolmogorov complexity of the multiplex and of the corresponding aggregated graph. We start by encoding the unweighted multiplex network $\mathcal{M}$ into the $N \times N$ *prime-weight matrix* Ω defined as follows:

$$\Omega_{ij} = \begin{cases} \displaystyle\prod_{\alpha: a_{ij}^{[\alpha]}=1} p^{[\alpha]} \\ 0 \qquad \text{if } a_{ij}^{[\alpha]}=0 \ \forall \ \alpha = 1, ..., M. \end{cases} \qquad (1)$$

The prime-weight matrix is obtained by assigning a distinct prime number $p^{[\alpha]}$ to each of the M layers of the multiplex, and then setting each element Ω_{ij} equal to the product of the primes associated to the layers where an edge between node i and node j actually exists. This procedure can be easily generalized to the case of weighted multiplex networks with integer weights, as explained in the Appendix A. Note that, given a certain assignment of prime numbers to the M layers, the matrix Ω is uniquely determined. Moreover, thanks to the unique factorization theorem, the prime-weight matrix preserves full information about the multiplex network $\mathcal{M}$, i.e., about the placement of all its edges.

A. Complexity of multiplex networks

We define the complexity of a multiplex network $\mathcal{M}$ with N nodes and M layers as the ratio

$$\mathcal{C}(\mathcal{M}) = \frac{KC(\Omega)}{KC(W)}, \qquad (2)$$
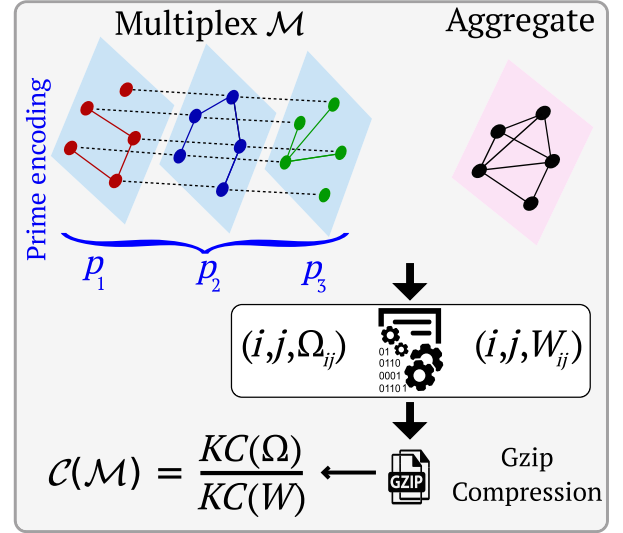

FIG. 1. The complexity $\mathcal{C}(\mathcal{M})$ of a multiplex network $\mathcal{M}$ is defined as the ratio between its Kolmogorov complexity and the Kolmogorov complexity of the associated single-layer aggregated graph. The multiplex is transformed into a string of bits by means of the prime-weight matrix. Since Kolmogorov complexity is not computable, we rely on an upper bound based on the size of the compressed string of bits associated to each object. A common way to obtain an upper bound is by computing the length of the string compressed through the Gzip algorithm.

where the numerator is the Kolmogorov complexity [36] of $\mathcal{M}$ and the denominator is the Kolmogorov complexity of the weighted aggregated graph associated to $\mathcal{M}$. In particular, the matrix Ω is the prime-weight matrix representation of $\mathcal{M}$, while W is the single-layer network obtained by aggregating all the M layers. We compute an approximation of the Kolmogorov complexity of a matrix by looking at the size of the compressed weighted edge list (see Appendix B and Supplemental Material Sec. S-1.5 [37] for details). The measure of complexity in Eq. (2) effectively quantifies the relative amount of additional algorithmic information needed to encode the multiplex network with respect to the amount needed to encode the corresponding single-layer aggregated graph (see Fig. 1 for a schematic representation of the procedure). As a particular case, $\mathcal{C}(\mathcal{M}) = 1$ if the multiplex network consists of M identical layers, but in general $\mathcal{C}(\mathcal{M}) \geq 1$, since the different possible arrangements of edges across the layers require more than one symbol to be encoded. The main hypothesis is that the higher the value of $\mathcal{C}(\mathcal{M})$, the larger the amount of information lost when representing the multiplex as a single-layer graph. In practice, if $\mathcal{C}(\mathcal{M}) \approx 1$, it would not make much difference to represent the multiplex as a single-layer graph, since the multiplex representation is not adding much more information. Conversely, when $\mathcal{C}(\mathcal{M}) > 1$ the aggregation of the multiplex into a single-layer graph would discard relevant information, and the larger the value of $\mathcal{C}(\mathcal{M})$, the more important it is to retain the full multiplex model. The code

for computing the complexity of a multiplex network is available in Ref. [38].

B. Synthetic multiplex networks

The fundamental ingredients contributing to the complexity of a multiplex network $\mathcal{M}$ as quantified by $\mathcal{C}(\mathcal{M})$ are the number of distinct pairs of nodes connected by an edge and the actual number of distinct symbols present in the prime-weight matrix Ω. In fact, both a larger number of connected pairs of nodes and a larger number of distinct symbols will in general result in a larger encoding and a larger value of $\mathcal{C}(\mathcal{M})$. Indeed, the number of symbols present in Ω is equal to the number of different multiplex motifs with two nodes [21].

The structural edge overlap o is a an easy-to-compute proxy for the variety of different multiedge configurations in a multiplex network (see Appendix C). In order to understand the effect of edge overlap on $\mathcal{C}(\mathcal{M})$, we considered ensembles of synthetic multiplex networks with different number of layers, where the total number of nodes and the average node degree on each layer are kept fixed ($N = 10\,000$, $\langle k \rangle = 6$), while the structural edge overlap o is tuned as explained in Appendix D. The results are shown in Fig. 2(a). As expected, $\mathcal{C}(\mathcal{M}) = 1$ in multiplex networks with M identical layers ($o = 1$). Indeed, when we start rewiring the edges of a multiplex with M identical layers, thus reducing the value of structural overlap, we expect $\mathcal{C}(\mathcal{M})$ to increase, since the prime-weight matrix contains a larger number of symbols. Conversely, when $o \approx 0$ each edge exists on exactly one of the M layers, meaning that the number of distinct edges in the multiplex is roughly equal to the number of distinct edges in the aggregated graph. As a consequence, we expect the values of their Kolmogorov complexity to not differ too much, and the corresponding value of $\mathcal{C}(\mathcal{M})$ to be somehow close to 1. This is exactly what we observe in Fig. 2(a), respectively for $o \approx 1$ and for $o \approx 0$. However, the most interesting result is that $\mathcal{C}(\mathcal{M})$ is a nonmonotonic function of the structural edge overlap, for any value of M. In particular, it is evident from Fig. 2(a) that $\mathcal{C}(\mathcal{M})$ always has a maximum for $o \in [0.15, 0.4]$, indicating that there exists indeed a maximum amount of additional information that a multiplex can encode with respect to the corresponding aggregate graph. We find it quite remarkable that the range at which $\mathcal{C}(\mathcal{M})$ peaks is compatible with the typical values of structural edge overlap observed in many real systems [30].

Structural hysteresis.—In order to fully explore the behavior of the complexity $\mathcal{C}$, we considered an ensemble of synthetic multiplex networks where we iteratively decrease and increase the structural overlap o (see Appendix D for details). Interestingly, we found two robust and distinct trajectories when the structural overlap is decreased (increased), characterized by a hysteresis loop [Fig. 2(b)]. In the simulations, we start from a multiplex network with $N = 10\,000$ nodes and $\langle k \rangle = 6$, where the
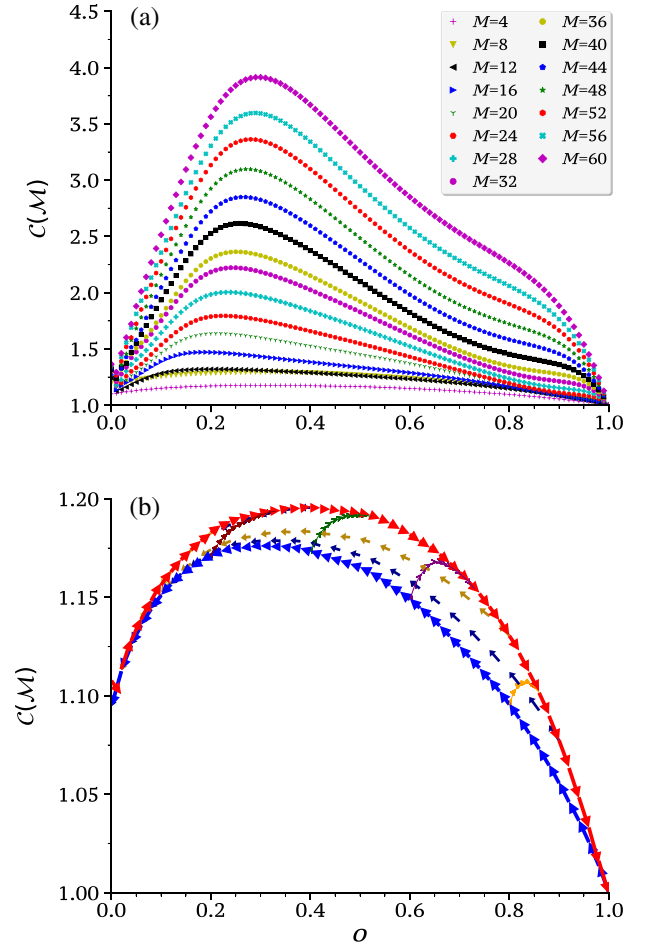

FIG. 2. (a) Complexity $\mathcal{C}(\mathcal{M})$ of ensembles of synthetic multiplex networks with a variable number of layers M and tunable structural overlap o. Each layer is an Erdös-Rényi graph with $N = 10\,000$ nodes and average degree equal to $\langle k \rangle = 6$. Irrespective of the number of layers M, the complexity $\mathcal{C}(\mathcal{M})$ always has a maximum for values of structural overlap in the interval [0.15, 0.4], which are compatible with the typical values of overlap measured in real-world networks [30] (see Supplemental Material Sec. S-1.3 for additional results on different synthetic networks [37]). (b) Hysteresis loop of $\mathcal{C}(\mathcal{M})$ for a multiplex with $N = 10\,000$ nodes, with $M = 4$ layers having each average degree $\langle k \rangle = 6$. Red arrows represent the trajectory observed when the structural edge overlap is reduced, while the blue arrows indicate the trajectories observed when the structural edge overlap is increased.

$M = 4$ layers are identical Erdös-Rényi random graphs, and we iteratively rewire the links in order to decrease the total edge overlap until we obtain a multiplex network with $o = 0$. After that, we successively increase the structural overlap until the system results in a multiplex network with $o = 1$. Remarkably, $\mathcal{C}(\mathcal{M})$ remains a nonmonotonic function of the structural edge overlap, but the plot of $\mathcal{C}(\mathcal{M})$ as a function of o reveals the presence of a structural hysteresis; i.e., the trajectory leading from $o = 1$ to $o = 0$ is different from the one obtained when the structural overlap is

increased from $o = 0$ to $o = 1$. This indicates that the procedures used to decrease and increase edge overlap are not ergodic, due to the intrinsic difference between the way overlap is created and destroyed. Indeed, if we start from a multiplex $\mathcal{M}$ with identical layers, the total number of ways in which a random rewiring can reduce the overlap is significantly larger than the total number of ways in which the edge overlap can be increased through a random move. Similar results are found when the graph on each layer is a regular or scale-free degree distribution, as shown in Supplemental Material, Fig. S-3 [37]. This result indicates that some caution is required when rewiring the edges of a multilayer graph, which is a problem we will explore in a future work.

C. Multiplex complexity and reducibility

One of the main issues of multidimensional datasets is that they normally contain redundant information. Consequently, the direct transformation of each type of relation available in a multidimensional dataset into a distinct layer of a multiplex network will possibly result in a structurally redundant representation of the original system. However, dealing with parsimonious models is always desirable, and is especially important in the case of multidimensional systems, where additional model complexity usually yields additional computational costs and raises some questions about the interpretability of the results. The "multiplex reducibility problem," originally formulated in Ref. [31], is the problem of finding low-dimensional representations of a multiplex network which preserve as much structural information as possible about the original system.

The multiplex complexity $\mathcal{C}(\mathcal{M})$ that we have defined provides a natural and meaningful way of obtaining reduced (low-dimensional) versions of a multiplex network over M layers. If we start from the original multiplex network $\mathcal{M}$ and we aggregate some of its layers, we obtain a reduced multiplex network $\mathcal{X}$ with $X \leq M$ layers, which will have a multiplex complexity $\mathcal{C}(\mathcal{X})$. We propose to quantify the normalized information content of the reduced multiplex network $\mathcal{X}$ as

$$q(\mathcal{X}) = \frac{\mathcal{C}(\mathcal{X})}{\log K_{\mathcal{X}}}, \tag{3}$$

where $K_{\mathcal{X}}$ is the number of distinct links in the multiplex $\mathcal{X}$. The normalization by $\log K_{\mathcal{X}}$ is necessary, since in general the Kolmogorov complexity of a bit string of length n is not smaller than $c + \log n$, for some $c \geq 0$ [39]. The length of the bit string associated to a multiplex network is proportional to the number of distinct links in the multiplex, hence, on average, a multiplex with a larger number of edges is expected to have a higher multiplex complexity. This is an inconvenience for the multiplex reducibility problem, since a network with a larger number of layers would in principle have a larger number of edges as well. The normalization used in $q(\mathcal{X})$ allows us to safely compare alternative low-dimensional multiplex networks obtained from the same original system.

Note that $q(\mathcal{X})$ behaves as a quality function, meaning that larger values of $q(\mathcal{X})$ indicate that the (possibly reduced) multilayer network $\mathcal{X}$ encodes a relatively larger amount of information with respect to the corresponding weighted aggregated graph $W_{\mathcal{X}}$. Hence, our goal is to find $\operatorname{argmax}[q(\mathcal{X})]$, which represents the optimally reduced
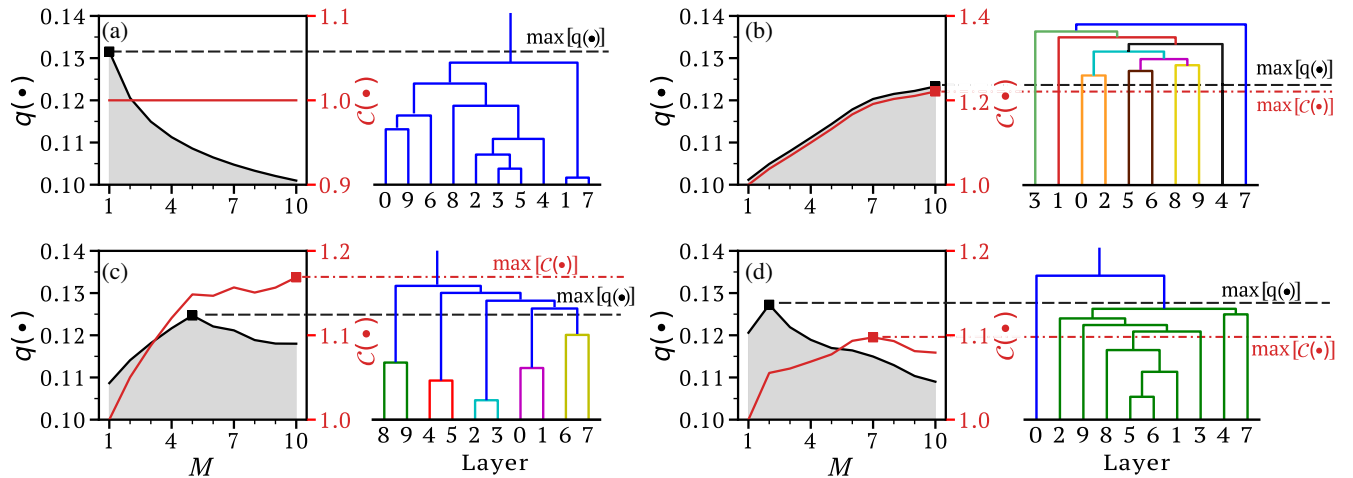

FIG. 3.   Reducibility of four synthetic multiplex networks with $N = 1000$ nodes, $\langle k \rangle = 4$, and $M = 10$ layers. The four panels correspond to (a) ten identical layers, (b) ten distinct layers, (c) five pairs of identical layers, (d) nine identical layers and one distinct layer. Note that all the distinct layers are independent realizations of Erdös-Rényi random graphs. For each synthetic multiplex we show both the global quality function $q(\bullet)$ and the complexity $\mathcal{C}(\bullet)$ as a function of the number of layers M (left-hand panel) and the dendrogram resulting from the greedy aggregation steps of the reducibility procedure (right-hand panel). Note that the maximum of the quality function q corresponds to the true partition of layers, while the complexity $\mathcal{C}$ generally fails to identify the correct partition.

multiplex $\mathcal{X}_{max}$ yielding the maximum value of information with respect to the aggregated graph. In particular, if all the layers of the multiplex network $\mathcal{X}$ are identical, then the maximum of q will always be at $1 = X \leq M$ layers, since the multiplex network and the aggregate graph are equivalent.

Maximizing $q(\mathcal{X})$ by enumerating all the possible partitions of the M layers is not feasible, since that number increases superexponentially with M. Hence, here we employed a classical agglomerative greedy algorithm to approximate the optimal solution (see Appendix E). The code for obtaining lower-dimensional representations of a multiplex network using $q(\bullet)$ is available in Ref. [38].

We started by testing the algorithm on *ad hoc* synthetic multiplex networks where some of the M layers are identical, thus reducing the number of truly distinct layers by construction. In Fig. 3 we report the results of the greedy reduction on four different synthetic benchmarks. In particular, we plot the global quality function $q(\bullet)$ and the complexity $\mathcal{C}(\bullet)$ as a function of the number of layers M (left-hand panels) and the dendrogram corresponding to the greedy aggregation steps (right-hand panels). In all the cases considered, the maximum of the quality function $q(\bullet)$ correctly identifies the partition made of truly distinct layers, while in general the complexity $\mathcal{C}(\bullet)$ fails to identify the correct partition (see Supplemental Material Sec. S-2 for results on a wide set of synthetic benchmarks [37]). This confirms our intuition that, by taking into account differences in the total number of edges of the multiplex, the quality function $q(\bullet)$ does a better job at discriminating between essential and redundant information. Indeed, while the multiplex with $M = 10$ identical layers is always aggregated into a single-layer graph [Fig. 3(a)], in the multiplex with all distinct layers the maximum of $q(\bullet)$ is attained by the initial configuration with ten layers [Fig. 3(b)].

TABLE I.   Reducibility of technological, social, and biological multiplex networks. From left to right, the columns report the number of layers in the original system (M), the structural edge overlap (o), the number of optimal layers [$M_{opt}(C)$] obtained when maximizing the quality function $q(\bullet)$, the value max [$q(\bullet)$], and the optimal value of complexity C_{opt} observed. The last two columns show the optimal number of layers $M_{opt}(VN)$ and the corresponding value of the quality function $q(\bullet)_{VN}$ obtained when using the multiplex structural reducibility procedure introduced in Ref. [31]. Although the two methods yield different results, they share similar features; i.e., technological multiplex networks are less likely to be reduced compared to biological and social systems.

Multiplex	M	o	$M_{opt}(C)$	max$[q(\bullet)_C]$	C_{opt}	$M_{opt}(VN)$	$q(\bullet)_{VN}$
London Tube [31]	13	0.006 810	11	0.183	1.125	2	0.499
Barcelona Tube	11	0.002 367	11	0.224	1.152	11	0.513
Bejing Tube	17	0.000 197	15	0.199	1.140	17	0.528
Berlin Tube	9	0.001 359	8	0.214	1.110	9	0.461
Airports North America [40]	143	0.003 958	129	0.143	1.271	93	0.697
Airports Europe [40]	175	0.003 185	163	0.162	1.413	109	0.675
Airports Asia [40]	213	0.005 477	209	0.180	1.636	146	0.291
Airports South America [40]	58	0.014 244	53	0.187	1.325	41	0.682
Airports Oceania [40]	37	0.014 532	27	0.185	1.192	31	0.665
Airports Africa [40]	84	0.006 876	74	0.191	1.274	65	0.719
EU airlines [18]	37	0.005 964	37	0.151	1.233	37	0.411
Train UK [41]	41	0.002 687	24	0.120	1.019	15	0.225
APS countries	10	0.451 138	10	0.176	1.618	2	0.047
Aarhus network [42]	5	0.189 093	5	0.201	1.291	2	0.158
Terrorist network [40]	4	0.153 558	4	0.171	1.166	2	0.239
Pierre Auger collaboration [43]	16	0.006 901	10	0.117	1.018	15	0.423
Arabidopsis [31]	7	0.007 690	6	0.105	1.023	7	0.421
Candida [31]	7	0.007 892	5	0.177	1.030	3	0.620
Celegans [31]	6	0.003 095	6	0.114	1.023	5	0.430
Drosophila [31]	7	0.004 389	5	0.098	1.011	6	0.379
Gallus [31]	6	0.012 923	5	0.179	1.043	5	0.577
Human Herpes-4 [31]	4	0.042 056	2	0.196	1.063	4	0.353
Human HIV-1 [31]	5	0.022 294	5	0.150	1.073	4	0.353
Mus [31]	7	0.010 776	7	0.106	1.041	6	0.375
Oryctolagus [31]	3	0.019 231	3	0.209	1.026	2	0.500
Plasmodium [31]	3	0.000 206	2	0.128	0.987	3	0.611
Rattus [31]	6	0.012 401	6	0.126	1.040	5	0.472
S. cerevisiae [31]	7	0.017 603	5	0.092	1.122	3	0.135
S. pombe [31]	7	0.007 070	5	0.099	1.067	2	0.206
Xenopus [31]	5	0.025 692	5	0.169	1.071	4	0.410

After having checked that $q(\bullet)$ identifies meaningful layer partitions in synthetic multiplex networks, we extended our analysis to real-world multiplex datasets. The results are reported in Table I. Note that most of the technological and biological multiplex networks in Table I admit reduced representations which have only a slightly smaller number of layers than the original systems. This is in agreement with the observation that in technological systems structural redundancy is purposely avoided. Similarly, the poor redundancy observed in biological multiplex networks is in line with the functionally different role played by each layer (protein interaction, functional dependence, mechanical interaction, and so on). However, technological systems exhibit consistently larger values of multiplex complexity than biological systems. A comparison with the reducibility algorithm proposed in Ref. [31] shows that our definition of multiplex complexity is in general more conservative, and often yields an optimal partition that has a slightly larger number of layers.

D. Structural and dynamical properties of reduced multiplex networks

It is important to note that a layer reduction procedure is expected to remove redundancies while maintaining as much information as possible about the original system. However, there is in general no *a priori* guarantee that the reduced multiplex obtained by aggregating some of the layers actually preserves any of the structural or dynamical properties of the original multiplex network to a given level of accuracy. To explore this aspect of layer reduction, we compared the distributions of four structural indicators in the original multiplex networks and in the networks obtained by using the aggregation procedure described above. In Fig. 4, we report the Kendall's τ correlation coefficient of the rankings induced by total node degree, node activity, participation coefficient, and node interdependence in both synthetic and real multiplex networks (see Appendix F for a formal description of those structural measures). Interestingly, in almost all the multiplex networks considered, the optimal partition identified by the multiplex complexity preserves most of the structural properties of the original system, as confirmed by the relatively high values of correlation ($\tau > 0.8$). Conversely, the optimal aggregations based on von Neumann entropy [31] often correspond to relatively lower values of correlation. We argue that this is a very desirable feature of the definition of complexity we have proposed. Indeed, a
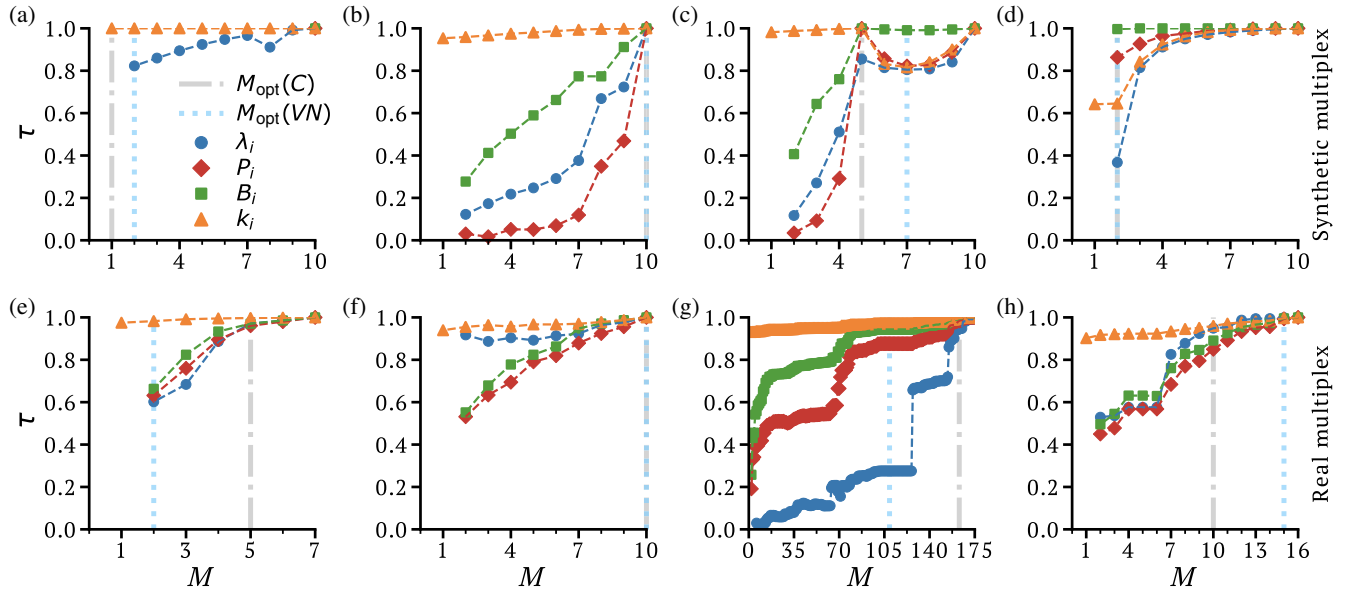

FIG. 4.   Kendall's τ correlation between the rankings induced by four structural descriptors in synthetic and real multiplex networks and the corresponding reduced networks with M layers. Panels (a)–(d) are for the same synthetic benchmarks shown in Fig. 3. The panels on the bottom row correspond to real-world systems, namely, (e) *Saccharomyces pombe*, (f) APS scientific collaborations among countries, (g) European airports, and (h) Pierre Auger scientific collaboration. The multiplex structural descriptors considered here are total node degree (k_i), node activity (B_i), participation coefficient (P_i), and node interdependence (λ_i). For each network we also indicate the number of layers in the optimal reduced system identified by multiplex complexity [$M_{\mathrm{opt}}(C)$, gray dash-dotted line] and by von Neumann entropy [$M_{\mathrm{opt}}(VN)$, blue dotted line] [31]. In all the synthetic cases, the quality function $q(\bullet)$ in Eq. (3) correctly identifies the optimal partition as the one where identical layers are aggregated, thus preserving the structural properties of the original multiplex. Conversely, in some cases the partitions found using the von Neumann entropy are different from the expected ones. Note that for the benchmark made of only identical layers (a), the correlation coefficient is not defined for two structural descriptors (constant behavior). For all the real-world networks the multiplex complexity provides a more conservative representation of the system and retains most of the properties of the original graph ($\tau > 0.8$ for all the structural descriptors), while the reduction based on von Neumann entropy might probably discard important information and/or create structural artifacts.

decrease of structural correlation is a clear indication that aggregation is creating structural artifacts. At the same time, the fact that the configuration found by using $q(\bullet)$ always yields high values of correlation with the original multiplex network confirms that the procedure is removing only truly redundant information, preserving most of the salient properties of the system. This is clearly visible when considering synthetic benchmarks [Figs. 4(a)–4(d)], where the optimal partition of the multiplex made of only distinct layers is known by construction. Despite the fact that the value of the Kendall's correlation decreases as the number of layers diminishes, the method based on multiplex complexity correctly identifies the optimal partition in all the cases considered.

Although the presence of high correlation between the structural properties of a multiplex network and its reduced counterpart indicates that the two systems are structurally similar, this will not guarantee in general that a dynamical process happening on the reduced multiplex network will exhibit a phenomenology similar to that observed on the original multiplex. As an example, we considered a multiplex susceptible-infected-susceptible (SIS) epidemic model [44] and we computed the epidemic threshold λ_c of the system at each step of the greedy aggregation procedure. In this process, the epidemic threshold depends on the contagion parameter γ/β that represents the ratio of intralayer versus interlayer contagion. In Fig. 5 we report the results of our analysis for both synthetic and real-world multiplex networks for different values of the contagion parameter γ/β. Note that any drop in the value of the critical

threshold corresponds to an important change in the structure of the reduced multiplex, e.g., to the formation of new (possibly artificial) structural patterns. This is easily observable in synthetic benchmarks [Figs. 5(a)–5(d)], where the epidemic threshold of the reduced multiplex remains the same as that of the original multiplex up to the point where $q(\bullet)$ is optimal, and then decreases abruptly. This means that, with respect to epidemic spreading, the reduced multiplex obtained by optimizing the quality function $q(\bullet)$ has basically the same dynamical behavior as the original multiplex, while further aggregations yield a system with different dynamics. As a consequence, the optimal reduced multiplex network (which has a smaller number of layers) can be used to make meaningful predictions about the dynamics of spreading of the original system. These results provide further evidence that the multiplex reduction based on Kolmogorov complexity somehow outperforms the reduction based on von Neumann entropy. See Figs. S-6 and S-7 in the Supplemental Material for additional evidence on several other synthetic benchmarks [37].

We obtain a similar but more intriguing picture for real-world multiplex networks, as shown in Figs. 5(e)–5(h). We notice that both methods preserve most of the information of the original multiplex, yielding approximately the same performance. Nevertheless, by looking at both the structural and dynamical features over the aggregation steps, it appears that the method based on $\mathcal{C}(\mathcal{M})$ is a bit more conservative and finds a reduced multiplex that simultaneously preserves as much as possible of both the structural and dynamical features of the original system. Indeed, the
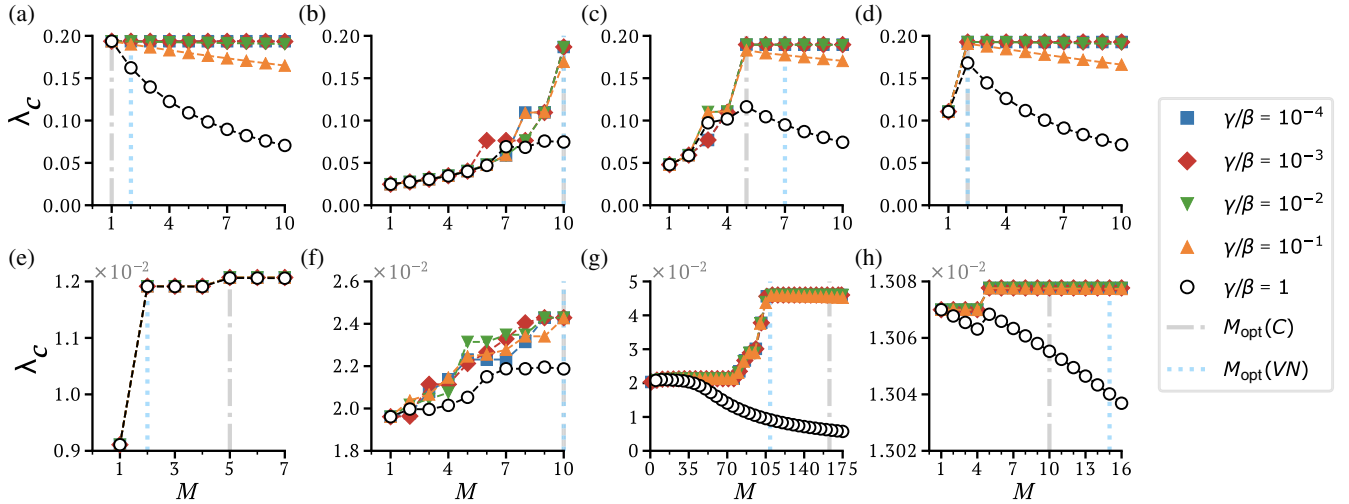

FIG. 5. Epidemic threshold λ_c for the multiplex SIS dynamics [44] as a function of the number of layers M and for different values of the contagion parameter γ/β, on the same synthetic and real multiplex networks shown in Fig. 4. For each network we indicate the number of layers in the optimal reduced system identified by multiplex complexity [$M_{\mathrm{opt}}(C)$, gray dash-dotted line] and by von Neumann entropy [$M_{\mathrm{opt}}(VN)$, blue dotted line] [31]. Interestingly, drops in the epidemic threshold correspond to dramatic changes in the structure of the multiplex, e.g., the formation of new structural patterns. This is clearly observed when considering the synthetic benchmarks (a)–(d), where the drops occur when truly distinct layers are aggregated. The reducibility procedures based on von Neumann entropy and multiplex complexity provide similar results in real-world networks [(e)–(h)], even though they yield different optimal partitions.

best layer partition identified by the method proposed here has high values of Kendall's correlation of structural properties ($\tau \approx 0.8$) and small variation of the epidemic threshold.

E. Complexity of time-varying multiplex networks

The multiplex complexity $\mathcal{C}(\mathcal{M})$ can be used to track the temporal evolution of the structure of time-varying multiplex networks. In Fig. 6 we show how the complexity of five large-scale multiplex networks has changed over time (see Appendix G for details about the datasets). Interestingly, $\mathcal{C}(\mathcal{M})$ provides a very good picture of the alternating behavior of the IMDb movie co-starring network over about a century [Fig. 6(a)] and of the network of thirty-five major assets in the NYSE and NASDAQ financial markets in the period 1998–2013 [Fig. 6(b)]. In both cases, local maxima of complexity are consistent with

the most notable periods of crisis in each dataset, while local minima of complexity seem to be precursors of renaissance in Internet Movie Database (IMDb) and of stability in the financial market. The value of complexity of scientific collaboration networks [American Physical Society (APS) and Web of Science, Fig. 6(c)] has remained stable around $\mathcal{C}(\mathcal{M}) = 1$ over the past thirty five years. This is mainly due to the fact that in these multiplex networks each layer represents a different field or subfield of science, and authors normally tend to publish in one or at most a couple of fields. In fact, the structural overlap of those multiplexes is always very small, and the large majority of pairs of nodes are connected in at most two layers. As a result, there is not indeed much benefit in considering the multiplex representation, since the information encoded in the different layers is comparable to that contained in the corresponding aggregated graph.
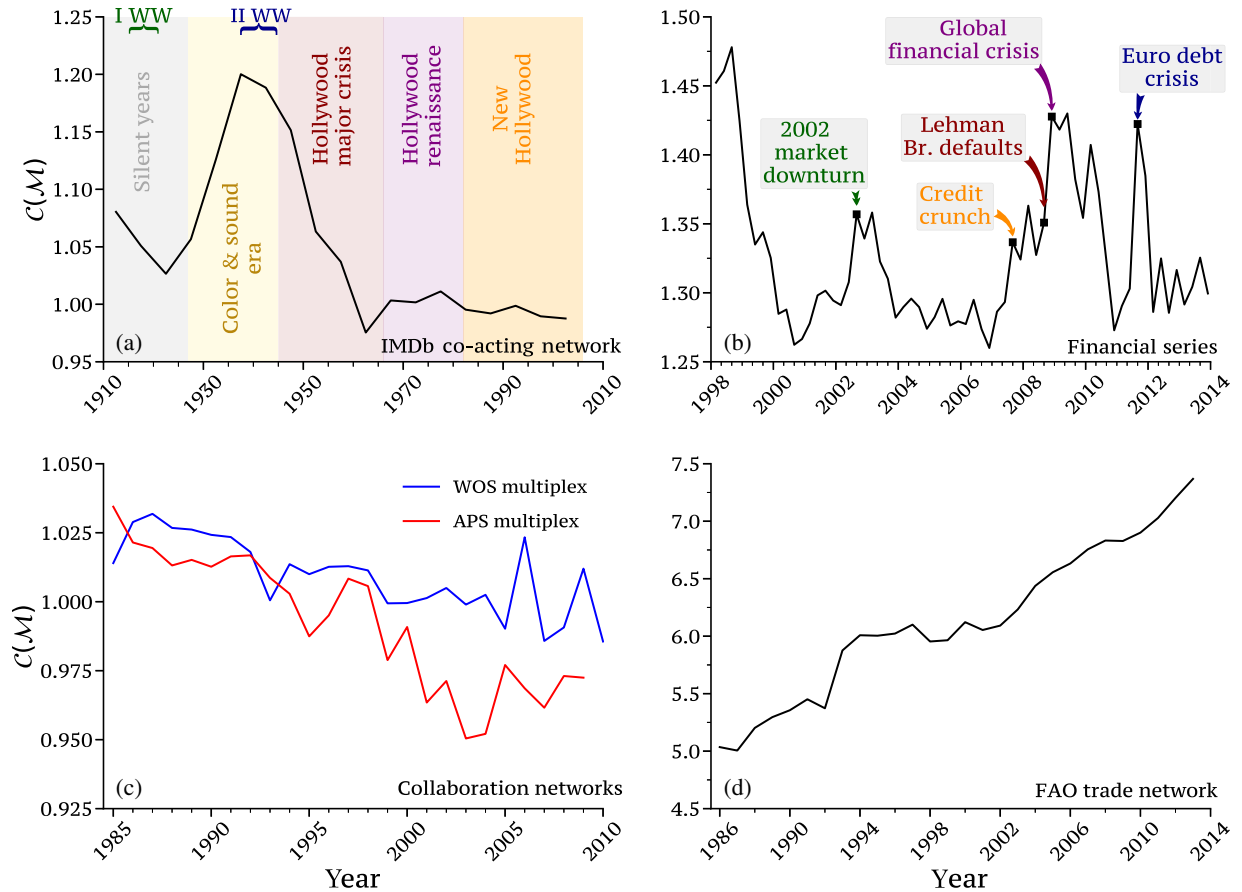

FIG. 6. Multiplex complexity as a function of time for four different time-varying multiplex networks, namely, (a) the IMDb co-starring network, (b) the financial multiplex constructed from price time series of thirty-five major assets in NYSE and NASDAQ, (c) the coauthorship multiplex of collaboration in American Physical Society (APS) journals and Web of Science (WOS), and (d) the FAO food import-export multiplex network. Notably, the most pronounced peaks of the complexity function in (a) and (b) correspond to periods of instability and crisis. Conversely, the values of complexity in the physics collaboration multiplex, for both the APS and WOS datasets (c), have remained pretty stable over time, and reveal that those systems indeed benefit only marginally from a multilayer representation. Finally, in the FAO food import-export multiplex network the complexity has kept increasing considerably over time (d), reflecting the relevant role played by globalization in the past twenty years in reshaping the international food market. More details and additional results on these four datasets are reported in Supplemental Material Sec. S-3 [37]. Abbreviations: WW—World War; Lehman Br.—Lehman Brothers.
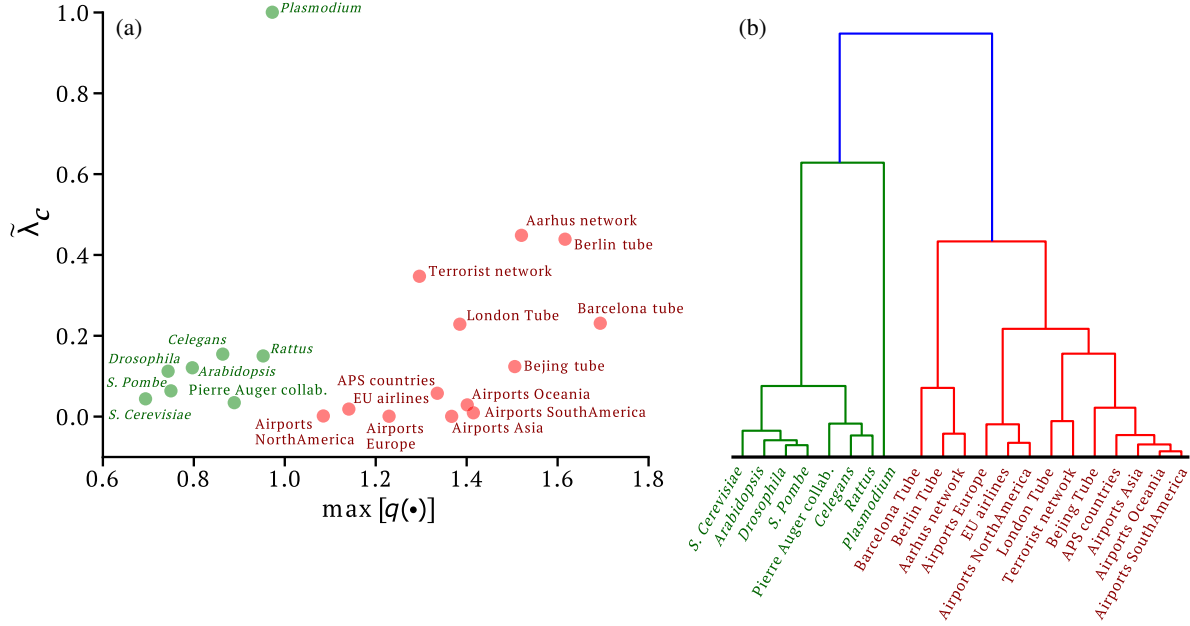
FIG. 7.   Multiplex cartography of real-world systems in the plane $\max[q(\bullet)] - \tilde{\lambda}_c$ (a) and the corresponding dendrogram obtained through Ward's hierarchical agglomerative clustering (b). It looks like these two structural descriptors alone are able to identify two large classes of real-world multiplex networks, namely, biological networks on one side (green cluster) and technological and social systems on the other side (red cluster).

It is worth noting that the complexity of the Food and Agriculture Organization of the United Nations (FAO) multiplex network of food exchange has kept increasing steadily in the past thirty years [Fig. 6(d)]. This is most probably linked to the globalization of commercial exchanges in general, which is reflected also in a more intricate pattern of relations among countries across a wide range of products. See Sec. S-3 of Supplemental Material and figures therein for further analyses of the same datasets [37].

F. Mapping multiplex networks

Finally, in Fig. 7 we show how multiplex complexity can be used to obtain a planar embedding of multiplex networks of different kind, and to reveal the presence of interesting clusters. For each multiplex network, we used the maximum value of the quality function $\max[q(\bullet)]$ as one of the coordinates and the normalized epidemic threshold $\tilde{\lambda}_c = \lambda_c/M_{\mathrm{opt}}(\mathcal{C})$ of the SIS dynamics [44] with contagion parameter $\gamma/\beta = 1$ as the other one. Note that $\tilde{\lambda}_c$ removes the dependence on the number of layers of the multiplex, making it possible to compare reduced multiplex networks with different numbers of layers. Moreover, since the epidemic threshold is intimately connected to the spreading dynamics on a graph, the information it provides is somehow orthogonal to that captured by multiplex complexity, which is instead a purely structural quantity. In Fig. 7(a) we indicated with different colors the two largest groups obtained through hierarchical clustering in the

$[\tilde{\lambda}_c, \max[q(\bullet)]]$ plane, while in Fig. 7(b) we show the corresponding dendrogram, highlighting all the aggregation steps, where at each step of the procedure we merge two clusters based on the minimum increase in total within-cluster variance over all possible pairs [45]. Interestingly, these two descriptors are already sufficient to cluster multiplex networks with different functions, so that all the biological multiplex networks appear in the same cluster and social and technological systems are put in another cluster. A more intriguing picture, where biological, social, and technological networks are put in three distinct clusters, is obtained when the normalized epidemic threshold is replaced by another dynamical descriptor, i.e., the maximal entropy rate per node [46] (results reported in Supplemental Material, Sec. S-4 [37]).

III. DISCUSSION

Quantifying the structural information encoded in a network is of fundamental importance to identify the key components of the system it represents. Indeed, information theory has already proven quite successful at extracting meaningful structural information from graphs [9,31,47] and at providing sound null models for different network-related tasks [48–50]. In the case of multidimensional datasets, and in particular of the multilayer networks constructed from them, assessing the actual amount of information contributed by each additional layer is of fundamental importance. Even though the current trends in data science seem to suggest otherwise, more data are

not always a bliss. Not all additional data available are indeed informative, and quite often more data implies more redundancy and more noise.

The algorithmic information approach proposed in this paper condenses the structural properties of a multiplex in a number—the multiplex complexity $\mathcal{C}(\mathcal{M})$. But since $\mathcal{C}(\mathcal{M})$ is defined as the ratio between the Kolmogorov complexity of the multiplex and that of the corresponding aggregated graph, its values allow us to assess to which extent a given multiplex representation of a system is more informative than a single-layer graph [e.g., if $\mathcal{C}(\mathcal{M})$ is larger than 1]. Consequently, it is meaningful to rank different multiplex networks according to their value of $\mathcal{C}(\mathcal{M})$, since those values are indeed telling us how much a multiplex representation deviates from the corresponding null-model hypothesis, i.e., that the system can be represented instead as a single-layer graph.

One of the most appealing aspects of $\mathcal{C}(\mathcal{M})$ is that it can be successfully employed to detect redundancy in a multiplex network and to obtain meaningful lower-dimensional representations of a system. In particular, it is worth noting that the quality function $q(\bullet)$ provides a consistent way to quantify the differences in complexity between alternative low-dimensional representations of the same system. The main difference between $\mathcal{C}(\mathcal{M})$ and $q(\bullet)$ is that the former quantifies the relative information encoded in a multiplex with respect to the aggregated graph, while the latter appropriately takes into account the fact that multiplex networks with different numbers of edges, even if obtained by reducing the same original multiplex network, will in general be associated to different aggregated graphs.

The results about multiplex reducibility shown in this paper have a double-pronged significance. On the one hand, the fact that almost all the multiplex networks analyzed admit a more compact lower-dimensional version is a warning against the quest to obtain more and more detailed data. There is a clear indication that not just more data points (edges), rather more informative data points are needed to complement the information already provided by existing layers. On the other hand, the possibility of reducing the number of layers of a multiplex has a lot of practical implications. Even simple multilayer structural descriptors, such as clustering coefficient, average shortest path, or any centrality measure based on paths scale superlinearly or exponentially as a function of the number of layers. Hence, a sound procedure to reduce the dimensionality of a network, without sacrificing information, would considerably speed up most of the computations on multiplex networks, without compromising on accuracy. Indeed, the optimal aggregations found by the quality function $q(\mathcal{M})$ often offer a substantial reduction in the number of layers needed to represent the system while still retaining most of the structural complexity of the original system (i.e., without introducing structural artifacts) as well as the salient characteristics determining the behavior of

dynamical processes happening on it (see the example of the preservation of the epidemic threshold in reduced multiplex networks). This is confirmed by synthetic benchmarks, where the optimal partitions is known, but also true for real systems, where $q(\mathcal{M})$ efficiently balances both the structural and dynamical features in the aggregation procedure. In a way, these results represent an important step toward finding minimal higher-order models that best compress information while preserving the original structural and dynamical patterns [51].

We believe it is quite remarkable that a reduction of layers based on the multiplex complexity proposed here usually produces reduced graphs that are sensibly different from those obtained using the classical reduction method based on von Neumann entropy [31]. We recall here that the definition of von Neumann entropy of a graph relies on a purely formal parallel between the quantum mixing operator and the rescaled Laplacian of the graph. It is true that the rescaled Laplacian is somehow related to the diffusive properties of the system, thus providing a concise description of the global properties of the graph. However, it is relatively difficult to pinpoint a specific graph property as responsible for a change of value of the von Neumann entropy. This fact was also noted by the authors of Ref. [31] in the Supplemental Material of the same paper, where they showed by simulation that a difference in the placement of a single edge of the graph can frequently result in relatively large fluctuations of the value of von Neumann entropy. Conversely, the multiplex complexity $\mathcal{C}(\mathcal{M})$ proposed here links quite closely to the traditional meaning of complexity of a system as the amount of information needed to fully describe it. This link is made possible by the prime-weight matrix, which encodes the full structure of the system in a string of bits. It is true that, in principle, the prime-weight matrix encoding depends on the chosen assignment of labels to nodes and primes to layers. However, it is remarkable that the assignment of primes to layers in increasing order of total number of edges provides a consistent approximation of Kolmogorov complexity, although a quite conservative one. Because of the way $\mathcal{C}(\mathcal{M})$ is defined, its value varies in a somehow predictable way if the edges of the graph are reorganized. In particular, if we add a single edge to an existing multiplex, then we can expect the value of multiplex complexity to change only slightly. Moreover, if the newly added edge increases the structural overlap of the multiplex, then the value of $\mathcal{C}(\mathcal{M})$ will increase if the original multiplex had a small structural overlap, or decrease if the multiplex had a large structural overlap. In this sense, $\mathcal{C}(\mathcal{M})$ is more closely associated to the structure of the system, and its changes provide information that are more readily interpretable.

Implementations of the algorithms to compute the complexity $\mathcal{C}(\mathcal{M})$ and to obtain reduced representations of a multiplex network based on the function $q(\bullet)$ are available in Ref. [38]. All the new datasets analyzed in the paper can be downloaded from Ref. [38].

APPENDIX A: PRIME-WEIGHT MATRIX ENCODING

An unweighted multiplex network $\mathcal{M}$ over N nodes is a set of M unweighted graphs (layers), each representing one type of interaction among the N nodes. In this framework, each node has a replica on each of the M layers, and the structure of each of the layers is in general distinct. The classical way to represent an unweighted multiplex network is by means of a vector of adjacency matrices $\mathcal{A} = \{a_{ij}^{[\alpha]}\}$, $\alpha = 1, \ldots, M$ [52]. The generic element $a_{ij}^{[\alpha]}$ of the adjacency matrix $A^{[\alpha]}$ at layer α is equal to 1 if and only if node i and node j are connected by a link at that layer, and zero otherwise. If we assign a distinct prime number $p^{[\alpha]}$ to each of the M layers, we can define the *prime-weight* matrix Ω whose elements are

$$\Omega_{ij} = \begin{cases} \prod_{\alpha : a_{ij}^{[\alpha]} = 1} p^{[\alpha]} & \\ 0 & \text{if } a_{ij}^{[\alpha]} = 0 \ \forall \ \alpha = 1, \ldots, M. \end{cases} \quad \text{(A1)}$$

The matrix $\Omega \in \mathbb{R}^{N \times N}$ is a compact encoding of the vector of adjacency matrices $\mathcal{A}$. In fact, thanks to the unique factorization theorem, the adjacency matrix of a generic layer α can be obtained from Ω by considering all the elements Ω_{ij} which are divisible by the corresponding prime $p^{[\alpha]}$. Note that this encoding works also for graphs with integer weights on the links, e.g., by associating to each pair of nodes (i, j) the number $\Omega_{ij} = \prod_{\alpha=1}^{M} (p^{[\alpha]})^{w_{ij}^{[\alpha]}}$, where $w_{ij}^{[\alpha]}$ is the weight of the edge (i, j) on layer α. Nevertheless, in this paper we always consider the case of unweighted multiplex networks; therefore, all the weights $w_{i,j}^{[\alpha]}$ are equal to 1.

Although the actual set of primes associated to the layers does not impact the construction of Ω, for practical reasons it makes sense to always use the sequence of the first M primes $\{2, 3, 5, \ldots\}$, since the actual number of bits required to store the matrix Ω is $O(N^2 M \log_2 \max_\alpha[p^{[\alpha]}])$. Note that, given a multilayer graph with M layers and a set of M distinct primes, we can construct $M!$ distinct prime-weight matrices, one for each of the possible permutations of the primes associated to the M layers. In this paper we choose a canonical prime association, that is the one that associates prime numbers to layers in increasing order

of their total number of edges. In practice, we assign the prime 2 to the layer with the smallest total number of edges $K^{[\alpha]} = 1/2 \sum_{ij} a_{ij}^{[\alpha]}$, the prime 3 to the layer with the second-smallest total number of edges, and so on (see Sec. S-1 of the Supplemental Material for a detailed discussion regarding the canonical prime association [37]).

APPENDIX B: MULTIPLEX COMPLEXITY

The Kolmogorov complexity $KC(S)$ of a bit string S is defined as the length of the shortest computer program that generates S as output [36]. However, it is easy to prove that $KC(S)$ is a noncomputable function [53]; thus it is only possible to approximate it. A common approach is to compress S using a given compression algorithm, and to consider the length of the compressed string S' as an estimate of $KC(S)$. In fact, it is possible to obtain S from the compressed string S' by using the decompression routine corresponding to the compression algorithm used to obtain S'. Thus, the concatenation of S' and the decompression routine is a program able to generate S, and its length is an upper bound for $KC(S)$. We associate a bit string $S(\mathcal{M})$ to a given multiplex network $\mathcal{M}$ by considering the bit string of the edge list associated to the prime-weight matrix Ω, where edges are listed in lexicographic order and each edge reports the corresponding entry of Ω (see Supplemental Material Sec. S-1.2 for a discussion about fluctuations due to node labeling [37]). We define the Kolmogorov complexity $KC(\mathcal{M})$ of the multiplex $\mathcal{M}$ as the length of the bit string $S'(\mathcal{M})$ obtained by compressing $S(\mathcal{M})$ with Gzip [54]. Note that this is not the only feasible choice, as any other compression algorithm can be used instead of Gzip for computing an upper bound of KC (see Supplemental Material Sec. S-1.4 and figures therein for additional comparisons between different compression algorithms [37]).

The complexity $\mathcal{C}(\mathcal{M})$ of a multiplex network $\mathcal{M}$ is equal to the Kolmogorov complexity of its prime-weight matrix Ω divided by the Kolmogorov complexity of the single-layer weighted matrix W, obtained by considering the aggregate binary matrix multiplied by the largest entry of Ω. In other terms, we express W as

$$W_{ij} = \begin{cases} \max_{i,j}[\Omega_{ij}] & \text{if } \Omega_{ij} \neq 0 \\ 0 & \text{otherwise.} \end{cases} \quad \text{(B1)}$$

Thus, the complexity $\mathcal{C}(\mathcal{M})$ of the multiplex $\mathcal{M}$ is defined as

$$\mathcal{C}(\mathcal{M}) = \frac{KC(\Omega)}{KC(W)}. \quad \text{(B2)}$$

Note that alternative representations of W are possible. One possibility is to set the weight of each existing edge $W_{ij} = o_{ij}$, where $o_{ij} = \sum_\alpha a_{ij}^{[\alpha]}$. Another option is to set

$W_{ij} = 2^{o_{ij}}$. Nevertheless, our definition of W is the only one that guarantees that $\mathcal{C}(\mathcal{M}) = 1$ when all the layers of the multiplex $\mathcal{M}$ are identical (see Supplemental Material Sec. S-1.5 and Fig. S-5 for comparisons between different representations of W [37]).

In general, the complexity of a multiplex might depend on the association of prime numbers to layers and on the actual node labeling. Numerical evidence confirms that the value of multiplex complexity obtained using the canonical prime association is always in the rightmost tail of the corresponding distribution (see Supplemental Material Sec. S-1.1 and table therein for further details [37]). As a consequence, the canonical prime association represents a conservative upper bound for the actual value of Kolmogorov complexity.

To reduce the effect of the other source of variability in the values of $\mathcal{C}(\mathcal{M})$ (i.e., the actual labeling of nodes, which affects the lexicographic ordering of the edge list), we define $\mathcal{C}(\mathcal{M})$ as the average of the multiplex complexity obtained by using the canonical prime association on 10^3 realizations of node relabeling on the same multiplex graph [see Supplemental Material S-1.2 for details on the distribution of $\mathcal{C}(\mathcal{M})$ as a function of node relabelings [37]].

APPENDIX C: STRUCTURAL EDGE OVERLAP

Given a multiplex $\mathcal{M}$ and a pair of nodes (i, j), the overlap o_{ij} of the pair is defined as the number of layers in which an edge exists between node i and node j. The matrix $O = \{o_{ij}\}$ is the overlapping matrix associated to the multiplex $\mathcal{M}$ [52]. The edge overlap o_s of a multiplex network is the expected number of layers in which a pair of nodes is connected by an edge [52,55]:

$$o_s = \frac{\sum_{i,j}^N o_{ij}}{M \sum_{i,j}^N \Theta(o_{ij})}, \tag{C1}$$

where $\Theta(x)$ is the Heaviside function, i.e., $\Theta(x) = 1$ if $x > 0$, and $1/M \le o_s \le 1$. In particular, $o_s = 1/M$ when there is no edge appearing in more than one layer, while $o_s = 1$ when all the M layers are identical. We define the structural overlap of a multiplex as

$$o = \frac{M}{M-1}\left(o_s - \frac{1}{M}\right), \tag{C2}$$

where the linear transformation $f(o_s) := [M/(M-1)] \times [o_s - (1/M)]$ maps o_s onto [0, 1].

APPENDIX D: SYNTHETIC NETWORKS

The results shown in Fig. 2(a) correspond to multiplex networks with $N = 10\,000$ nodes. The plots are obtained by starting from a multiplex network with M identical Erdös-Rényi random graphs as layers (thus having

structural overlap $o = 1$), and then iteratively rewiring the edges on each layer in order to decrease the structural overlap to $o = 0$. Edge rewiring is performed by selecting a pair of edges and swapping their end points uniformly at random. This rewiring procedure is similar to the one used in Ref. [30], and preserves the degree sequence at each layer. Consequently, the layers of all the multiplex networks obtained through relabeling are Erdös-Rényi random graphs belonging to the same ensemble. The value of multiplex complexity corresponding to a certain value of structural overlap is obtained by averaging over 10^2 distinct realizations.

The algorithm to increase the structural overlap of a multiplex is similar to the one used to decrease it. The only difference, however, is that to increase the structural overlap a rewiring is accepted only if it results in the increase of edge overlap of at least one of the two edges involved in the rewiring.

APPENDIX E: REDUCIBILITY

Computing the global maximum of the quality function $q(\bullet)$ is in general computationally unfeasible, since it requires us to enumerate all the possible partitions of M objects. This is an NP-hard problem that requires a number of operations that scales superexponentially with M [56]. In order to avoid this problem, we used instead a greedy algorithm, which reduces the time complexity to $\mathcal{O}(M^2)$. The algorithm starts from the original multiplex with M layers and at each step computes the complexity of the two-layer multiplex networks corresponding to all the possible pairs of layers. We call $\bar{D}$ the pair of layers with the maximum value of complexity, and we consider the set of pairs of layers whose overlap is larger than or equal to that of $\bar{D}$. Then, we aggregate the pair of layers D of that set yielding the smallest value of complexity $\mathcal{C}(D)$. Aggregation is performed by considering the union of the edges in the two layers. The rationale behind this choice is that if two layers form a duplex with relatively high overlap and small complexity, then they are similar enough and can be thus flattened in a single layer. The iteration of this procedure will result in a sequence of multiplex networks with $\{M, M-1, M-2, \ldots, 2, 1\}$ layers. Among those M reduced multiplex networks, we choose the one yielding the largest value of $q(\bullet)$.

APPENDIX F: STRUCTURAL MULTIPLEX MEASURES

To analyze the structural properties of both synthetic and real multiplex networks, we considered four different structural descriptors [40,52].

(1) Total degree:

$$k_i = \sum_\alpha k_i^{[\alpha]} = \sum_\alpha \sum_j a_{ij}^{[\alpha]},$$

i.e., the total number of links incident on node i across all the layers.

(2) Node participation coefficient:

$$P_i = \frac{M}{M-1}\left[1 - \sum_\alpha \left(\frac{k_i^{[\alpha]}}{k_i}\right)^2\right],$$

which measures the heterogeneity of the number of neighbours of node i across the layers.

(3) Node activity:

$$B_i = \sum_\alpha \Theta(k_i^{[\alpha]}),$$

i.e., the number of layers on which node i has at least one neighbor. Here, Θ represents the Heaviside step function.

(4) Node interdependence:

$$\lambda_i = \frac{1}{N-1}\sum_{\substack{j\in N \\ j\neq i}}\frac{\psi_{ij}}{\sigma_{ij}}.$$

In the expression, ψ_{ij} is the number of shortest paths between i and j that span across more than one layer, while σ_{ij} is the total number of shortest paths between i and j. If $\lambda_i \approx 1$, then i fully exploits the multiplex structure of the system to reach other nodes, while if $\lambda_i \approx 0$, node i reaches other nodes through shortest paths whose edges are on just one layer. In Supplemental Material Sec. S-5 we describe an algorithm to compute node interdependence that exploits the prime-weight matrix introduced in this paper [37].

APPENDIX G: MULTIPLEX DATASETS

The datasets introduced in this paper for the reducibility comparisons are (i) the undirected routes of the 11 lines of the Barcelona tube network [57], (ii) the 9 lines of the Berlin tube [58], (iii) the 17 lines of the Beijing subway [59], and (iv) the scientific collaboration among countries (APS countries) obtained considering the papers published in the journals of the American Physical Society. For the latter, starting with the multiplex dataset introduced in Ref. [40], we constructed a weighted multiplex collaboration network, in which nodes represent countries and a link connects two countries if scientists based in those countries coauthored a paper together. Authors having multiple affiliations were considered as belonging to multiple countries. The weight on each link represents the number of coauthorship relations between the corresponding two countries. In our analysis the unweighted version of such system has been used.

In addition, the time-varying datasets used in this paper are (i) the IMDb co-starring network [40], (ii) the financial multiplex network constructed from price time series of

35 major assets in NYSE and NASDAQ [60], (iii) the physics collaboration multiplex network of the American Physical Society (APS) and Web of Science (WOS) [40], and (iv) the FAO food import-export multiplex network [61]. From each original dataset [(i), (iii), (iv)], we constructed a time-varying multiplex network by partitioning the original system in temporal windows of one year. In this process, we associate to each time window the corresponding static multiplex network containing all the links registered in that year.

---

[1] M. Newman, *Networks: An Introduction* (Oxford University Press, New York, 2010).

[2] V. Latora, V. Nicosia, and G. Russo, *Complex Networks: Principles, Methods and Applications* (Cambridge University Press, Cambridge, England, 2017).

[3] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, *Epidemic Processes in Complex Networks*, Rev. Mod. Phys. **87**, 925 (2015).

[4] A. Arenas, A. Díaz-Guilera, J. Kurths, Y. Moreno, and C. Zhou, *Synchronization in Complex Networks*, Phys. Rep. **469**, 93 (2008).

[5] E. T. Jaynes, *Information Theory and Statistical Mechanics*, Phys. Rev. **106**, 620 (1957).

[6] G. Bianconi, *The Entropy of Randomized Network Ensembles*, Europhys. Lett. **81**, 28005 (2008).

[7] K. Anand and G. Bianconi, *Entropy Measures for Networks: Toward an Information Theory of Complex Topologies*, Phys. Rev. E **80**, 045102(R) (2009).

[8] M. Dehmer, *Information Processing in Complex Networks: Graph Entropy and Information Functionals*, Appl. Math. Comput. **201**, 82 (2008).

[9] F. Passerini and S. Severini, *Quantifying Complexity in Networks: The von Neumann Entropy*, Int. J. Agents Technol. Syst. **1**, 58 (2009).

[10] A. Mowshowitz and M. Dehmer, *Entropy and the Complexity of Graphs Revisited*, Entropy **14**, 559 (2012).

[11] M. Dehmer and A. Mowshowitz, *A History of Graph Entropy Measures*, Inf. Sci. **181**, 57 (2011).

[12] G. Cimini, T. Squartini, F. Saracco, D. Garlaschelli, A. Gabrielli, and G. Caldarelli, *The Statistical Physics of Real-World Networks*, Nat. Rev. Phys. **1**, 58 (2019).

[13] M. Morzy, T. Kajdanowicz, and P. Kazienko, *On Measuring the Complexity of Networks: Kolmogorov Complexity versus Entropy*, Complexity **2017**, 3250301 (2017).

[14] H. Zenil, N. Kiani, and J. Tegnér, *A Review of Graph and Network Complexity from an Algorithmic Information Perspective*, Entropy **20**, 551 (2018).

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

[16] S. Boccaletti, G. Bianconi, R. Criado, C. I. del Genio, J. Gómez-Gardeñes, M. Romance, I. S. Nadal, Z. Wang, and M. Zanin, *The Structure and Dynamics of Multilayer Networks*, Phys. Rep. **544**, 1 (2014).

[17] G. Bianconi, *Multilayer Networks: Structure and Function* (Oxford University Press, Oxford, 2018).

[18] A. Cardillo, J. Gómez-Gardeñes, M. Zanin, M. Romance, D. Papo, F. Pozo, and S. Boccaletti, *Emergence of Network Features from Multiplexity*, Sci. Rep. **3**, 1344 (2013).

[19] R. Gallotti, M. A. Porter, and M. Barthelemy, *Lost in Transportation: Information Measures and Cognitive Limits*, Sci. Adv. **2**, e1500445 (2016).

[20] M. De Domenico, S. Sasai, and A. Arenas, *Mapping Multiplex Hubs in Human Functional Brain Networks*, Front. Neurosci. **10**, 326 (2016).

[21] F. Battiston, V. Nicosia, M. Chavez, and V. Latora, *Multilayer Motif Analysis of Brain Networks*, Chaos **27**, 047404 (2017).

[22] S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, and S. Havlin, *Catastrophic Cascade of Failures in Interdependent Networks*, Nature (London) **464**, 1025 (2010).

[23] S. Gómez, A. Díaz-Guilera, J. Gómez-Gardeñes, C. J. Pérez-Vicente, Y. Moreno, and A. Arenas, *Diffusion Dynamics on Multiplex Networks*, Phys. Rev. Lett. **110**, 028701 (2013).

[24] V. Nicosia, P. S. Skardal, A. Arenas, and V. Latora, *Collective Phenomena Emerging from the Interactions between Dynamical Processes in Multiplex Networks*, Phys. Rev. Lett. **118**, 138302 (2017).

[25] D. Soriano-Paños, L. Lotero, A. Arenas, and J. Gómez-Gardeñes, *Spreading Processes in Multiplex Metapopulations Containing Different Mobility Networks*, Phys. Rev. X **8**, 031039 (2018).

[26] C. Granell, S. Gómez, and A. Arenas, *Dynamical Interplay between Awareness and Epidemic Spreading in Multiplex Networks*, Phys. Rev. Lett. **111**, 128701 (2013).

[27] J. P. Gleeson, K. P. O'Sullivan, R. A. Baños, and Y. Moreno, *Effects of Network Structure, Competition and Memory Time on Social Spreading Phenomena*, Phys. Rev. X **6**, 021019 (2016).

[28] M. De Domenico, C. Granell, M. A. Porter, and A. Arenas, *The Physics of Spreading Processes in Multilayer Networks*, Nat. Phys. **12**, 901 (2016).

[29] L. Lacasa, I. P. Mariño, J. Miguez, V. Nicosia, É. Roldán, A. Lisica, S. W. Grill, and J. Gómez-Gardeñes, *Multiplex Decomposition of Non-Markovian Dynamics and the Hidden Layer Reconstruction Problem*, Phys. Rev. X **8**, 031038 (2018).

[30] M. Diakonova, V. Nicosia, V. Latora, and M. San Miguel, *Irreducibility of Multilayer Network Dynamics: The Case of the Voter Model*, New J. Phys. **18**, 023010 (2016).

[31] M. De Domenico, V. Nicosia, A. Arenas, and V. Latora, *Structural Reducibility of Multilayer Networks*, Nat. Commun. **6**, 6864 (2015).

[32] J. Iacovacci, Z. Wu, and G. Bianconi, *Mesoscopic Structures Reveal the Network between the Layers of Multiplex Datasets Data Sets*, Phys. Rev. E **92**, 042806 (2015).

[33] T.-C. Kao and M. A. Porter, *Layer Communities in Multiplex Networks*, J. Stat. Phys. **173**, 1286 (2018).

[34] N. Stanley, S. Shai, D. Taylor, and P. J. Mucha, *Clustering Network Layers with the Strata Multilayer Stochastic Block Model*, IEEE Trans. Network Sci. Eng. **3**, 95 (2016).

[35] C. De Bacco, E. A. Power, D. B. Larremore, and C. Moore, *Community Detection, Link Prediction, and Layer Interdependence in Multilayer Networks*, Phys. Rev. E **95**, 042317 (2017).

[36] A. N. Kolmogorov, *On Tables of Random Numbers*, Theor. Comput. Sci. **207**, 387 (1998).

[37] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevX.10.021069 for in-depth analysis of multiplex complexity, additional results and analysis on synthetic and real-world multiplex networks, and comparisons with possible alternative formulations.

[38] https://github.com/andresantoro/ALCOREM.

[39] J.-P. Delahaye and H. Zenil, *Numerical Evaluation of Algorithmic Complexity for Short Strings: A Glance into the Innermost Structure of Randomness*, Appl. Math. Comput. **219**, 63 (2012).

[40] V. Nicosia and V. Latora, *Measuring and Modeling Correlations in Multiplex Networks*, Phys. Rev. E **92**, 032805 (2015).

[41] A. Santoro, V. Latora, G. Nicosia, and V. Nicosia, *Pareto Optimality in Multilayer Network Growth*, Phys. Rev. Lett. **121**, 128302 (2018).

[42] https://comunelab.fbk.eu/data.php.

[43] M. De Domenico, A. Lancichinetti, A. Arenas, and M. Rosvall, *Identifying Modular Flows on Multilayer Networks Reveals Highly Overlapping Organization in Social Systems*, Phys. Rev. X **5**, 011027 (2015).

[44] E. Cozzo, R. A. Baños, S. Meloni, and Y. Moreno, *Contact-Based Social Contagion in Multiplex Networks*, Phys. Rev. E **88**, 050801(R) (2013).

[45] J. H. Ward, Jr., *Hierarchical Grouping to Optimize an Objective Function*, J. Am. Stat. Assoc. **58**, 236 (1963).

[46] F. Battiston, V. Nicosia, and V. Latora, *Efficient Exploration of Multiplex Networks*, New J. Phys. **18**, 043035 (2016).

[47] M. Rosvall and C. T. Bergstrom, *Maps of Random Walks on Complex Networks Reveal Community Structure*, Proc. Natl. Acad. Sci. U.S.A. **105**, 1118 (2008).

[48] T. P. Peixoto, *Model Selection and Hypothesis Testing for Large-Scale Network Models with Overlapping Groups*, Phys. Rev. X **5**, 011033 (2015).

[49] T. P. Peixoto, *Reconstructing Networks with Unknown and Heterogeneous Errors*, Phys. Rev. X **8**, 041011 (2018).

[50] A. Godoy-Lorite, R. Guimerà, C. Moore, and M. Sales-Pardo, *Accurate and Scalable Social Recommendation Using Mixed-Membership Stochastic Block Models*, Proc. Natl Acad. Sci. U.S.A. **113**, 14207 (2016).

[51] R. Lambiotte, M. Rosvall, and I. Scholtes, *From Networks to Optimal Higher-Order Models of Complex Systems*, Nat. Phys. **15**, 313 (2019).

[52] F. Battiston, V. Nicosia, and V. Latora, *Structural Measures for Multiplex Networks*, Phys. Rev. E **89**, 032804 (2014).

[53] G. J. Chaitin, *On the Simplicity and Speed of Programs for Computing Infinite Sets of Natural Numbers*, J. ACM **16**, 407 (1969).

[54] https://www.gzip.org.

[55] F. Battiston, V. Nicosia, V. Latora, and M. San Miguel, *Layered Social Influence Promotes Multiculturality in the Axelrod Model*, Sci. Rep. **7**, 1809 (2017).

[56] E. T. Bell, *Exponential Numbers*, Am. Math. Mon. **41**, 411 (1934).

[57] https://www.tmb.cat/.

[58] https://www.berlin.de/en/public-transportation/.

[59] https://www.bjsubway.com/.

[60] N. Musmeci, V. Nicosia, T. Aste, T. Di Matteo, and V. Latora, *The Multiplex Dependency Structure of Financial Markets*, Complexity **2017**, 9586064 (2017).

[61] http://www.fao.org/statistics/databases.