



## Geometric detection of hierarchical backbones in real networks

Elisenda Ortiz <sup>1,2</sup>, Guillermo García-Pérez,<sup>3,4</sup> and M. Ángeles Serrano <sup>1,2,5,\*</sup>

<sup>1</sup>*Departament de Física de la Matèria Condensada, Universitat de Barcelona, Martí i Franquès 1, 08028 Barcelona, Spain*

<sup>2</sup>*Universitat de Barcelona Institute of Complex Systems (UBICS), Universitat de Barcelona, Barcelona, Spain*

<sup>3</sup>*QTF Centre of Excellence, Turku Centre for Quantum Physics, Department of Physics and Astronomy, University of Turku, FI-20014 Turun Yliopisto, Finland*

<sup>4</sup>*Complex Systems Research Group, Department of Mathematics and Statistics, University of Turku, FI-20014 Turun Yliopisto, Finland*

<sup>5</sup>*ICREA, Pg. Lluís Companys 23, E-08010 Barcelona, Spain*



(Received 29 June 2020; accepted 26 August 2020; published 29 September 2020)

Hierarchies permeate the structure of real networks, whose nodes can be ranked according to different features. However, networks are far from treelike structures and the detection of hierarchical ordering remains a challenge, hindered by the small-world property and the presence of a large number of cycles, in particular clustering. Here, we use geometric representations of undirected networks to achieve an enriched interpretation of hierarchy that integrates features defining the popularity of nodes and similarity between them, such that the more similar a node is to a less popular neighbor the higher the hierarchical load of the relationship. The geometric approach allows us to measure the local contribution of nodes and links to the hierarchy within a unified framework. Additionally, we propose a link filtering method, the *similarity filter*, able to extract hierarchical backbones containing the links that represent statistically significant deviations with respect to the maximum entropy null model for geometric heterogeneous networks. We applied our geometric approach to the detection of similarity backbones of real networks in different domains and found that the backbones preserve local topological features at all scales. Interestingly, we also found that similarity backbones favor cooperation in evolutionary dynamics modeling social dilemmas.

DOI: [10.1103/PhysRevResearch.2.033519](https://doi.org/10.1103/PhysRevResearch.2.033519)

### I. INTRODUCTION

Many real systems display a hierarchical organization [1] where higher status members dominate over lower-graded ones, according to a certain measure of power, wealth, importance or influence. Examples are ubiquitous in living systems, including molecular regulators governing gene expression [2], animal communities of eusocial insects [3], dominant-subordinate relationships in mammals [4], and different structures—companies, political parties, courts, military, organized religion, etc.—in human society [5]. Additionally, hierarchical organization can be found in nonliving systems such as computer generated imagery (CGI) [6], grammatical theory of language [7], or the structure of a musical composition [8]. Hierarchies are, thus, ubiquitous, and shape more easily controllable structures [9] that can emerge as the result of opposing forces, such as competition between individuals [10] or a combination of cooperation and imitation strategies [11].

Complex networked systems, however, present important challenges when it comes to the detection of hierarchies due to the lack of a unique and unambiguous stratification scheme,

possibly including nestedness or layered structure. This is caused by the small world property and the presence of a large number of cycles of different lengths in their topologies, in particular clustering [12], so that networks' organization deviates strongly from treelike. In directed networks, link directionality can be exploited to ease the problem and hierarchical order can be detected using, for instance, penalty-function minimization strategies [13–16]. Nevertheless, most frequently the only meaningful or available representation of a complex system is an undirected graph. Within this architecture, a hierarchy is typically defined as a ranking where the status of a node becomes determined by some heterogeneous topological property, for instance degree [17] or some other centrality measure [9,18]. However, other attributes shape as well the hierarchical structure of real networks, like similarity between nodes [19]. Clearly, the control exerted by a higher-status node over a lower status one will be more effective when there exists closeness or affinity between them. Conversely, the strength of hierarchical relations gets dissolved as nodes lose their proximity and become dissimilar.

Here, we integrate degree rank, or popularity, and similarity between nodes in an enriched interpretation of hierarchy, valid for real networks. For this purpose, we capture network architecture using geometric network maps [20,21] and models [19,22], which naturally encode popularity and similarity attributes of nodes as coordinates in an underlying metric space. Exploiting the geometric approach, we are able to characterize the individual contribution of each node and each link to the hierarchical structure of a network. Moreover,

\*marian.serrano@ub.edu

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

we exploit the great heterogeneity found in the hierarchy load of links to propose a filtering method, the similarity filter, that offers a practical procedure to extract a hierarchical similarity backbone of a network. The obtained similarity backbones contain the links that represent statistically relevant interactions with respect to the maximally random geometric network organization, constrained by node degrees and clustering [23]. The similarity filter preserves network features at all scales while pruning a large number of links, in the spirit of the disparity filter for weighted networks [24]. However, the similarity filter has a different purpose and operates on the basis of different models in a completely distinct framework, that of unweighted undirected networks. To illustrate the use and results of the similarity filter, we extracted and analyzed the hierarchical similarity backbones of several real-world networks from different domains. Finally, we explored the role of hierarchical similarity backbones in evolutionary dynamical processes, which historically have been argued to be sensitive to the hierarchical organization of complex architectures [25,26]. Implementing an evolutionary prisoner's dilemma game [27–30] on the real networks under study, we discovered that the similarity backbones tend to achieve final states of greater cooperation as compared to the corresponding original networks, when initial conditions are equivalent.

## II. HIERARCHY LOAD OF LINKS AND NODES

We base our definitions of hierarchy on geometric maps of real networks obtained from geometric network models [19,22]. In these models, the high clustering typically observed in real networks emerges naturally as a result of the metric properties of an underlying space, which abstracts the similarity space among nodes. In this paper, we use the  $\mathcal{S}^1$  model, in which the similarity space is a one-dimensional sphere (a circle). Nodes thus have an angular coordinate  $\theta_i$ , so that the more similar two nodes are the shorter is the angular separation,  $\Delta\theta_{ij} = \min(|\theta_i - \theta_j|, 2\pi - |\theta_i - \theta_j|)$ , between them. Nodes are also characterized by a hidden degree  $\kappa_i$ , directly related to the observed degree, standing for popularity or status. The probability that two nodes  $i$  and  $j$  connect is a function that increases with their similarity and with the product of their hidden degrees

$$p_{ij} = \frac{1}{1 + \left(\frac{R\Delta\theta_{ij}}{\mu\kappa_i\kappa_j}\right)^\beta}, \quad (1)$$

where  $R$  is the radius of the similarity circle, which we choose to be equal to  $N/(2\pi)$  for  $N$  nodes, so that the density of nodes in the circle is equal to one, and parameters  $\beta$  and  $\mu$  control the mean clustering coefficient and the average degree of the network, respectively.

The  $\mathcal{S}^1$  model is able to reproduce many of the features widely observed in real complex networks, such as scale-freeness, high levels of clustering and the small-world property, among others [19]. Interestingly, the  $\mathcal{S}^1$  model is the only model able to produce maximum entropy ensembles with power-law degree distributions and clustering and without nonstructural degree correlations [23]. Moreover, the model also allows us to construct geometric maps of real networks through a process called *network embedding*. Given a real

network, one can find the values  $\{\kappa_i, \theta_i\}$  for every node  $i$ , as well as the global parameters  $\mu$  and  $\beta$ , that maximize the likelihood for the observed network to be generated by the model [20]. The congruency of the  $\mathcal{S}^1$  model with real networks results in very meaningful embeddings, which have proven to be useful for network navigation [20,31,32], and for the discovery of symmetries in the structure of real networks such as self-similarity at different length scales [33]. Notice that the distribution of angular positions of nodes in the  $\mathcal{S}^1$  model is assumed to be homogeneous, while typically real network embeddings show that nodes form geometric communities in similarity space [20,34–36]. The geometric embeddings of real networks used in this work were computed using the mapping tool Mercator [21].

We use the purely geometric isomorphic version of the  $\mathcal{S}^1$  model, named the  $\mathcal{H}^2$  model [22], for visualization purposes. In the  $\mathcal{H}^2$  model, the hidden degree is transformed into a radial coordinate in a hyperbolic two-dimensional disk, such that higher degree nodes are placed closer to the centre of the disk. The angular coordinate remains as in the  $\mathcal{S}^1$  circle and the connection probability becomes a decreasing function of the hyperbolic distance between nodes, see Fig. 1(a). For further details refer to section Methods B.

### A. Definition of hierarchy load

We characterize the local contribution of a link or a node to the hierarchical structure of a network by measuring its *hierarchy load*, which depends on status, similarity, and the reference provided by a null model to discount the effects of random fluctuations. We use the  $\mathcal{S}^1$  as a null model since it is the maximum entropy model for geometric networks with heterogeneous degrees [23]. It provides expectations for the distribution of hierarchy loads in a pure random assignment of angular positions of nodes given a degree distribution and a level of clustering, so that anomalous fluctuations can be detected.

Given a geometric network embedding, where a node  $i$  has coordinates  $\{\kappa_i^{\text{obs}}, \theta_i^{\text{obs}}\}$ , we consider that popularity, corresponding to the hidden degree or, equivalently, to the radial position in the hyperbolic plane, is a measure of status. This means that a node  $i$  has a lower-status neighbor  $j$  when  $\kappa_j^{\text{obs}} < \kappa_i^{\text{obs}}$ . Similarity between the two nodes is represented by their angular separation in the network map,  $\Delta\theta_{ij}^{\text{obs}} = \min(|\theta_i^{\text{obs}} - \theta_j^{\text{obs}}|, 2\pi - |\theta_i^{\text{obs}} - \theta_j^{\text{obs}}|)$ .

First, we define the hierarchy load  $h_{ij}$  of a link between node  $i$  and its lower-status neighbor  $j$  as the probability of obtaining a similarity distance between them in the  $\mathcal{S}^1$  model greater than the one observed in the map

$$h_{ij} = P(\Delta\theta_{ij} > \Delta\theta_{ij}^{\text{obs}}). \quad (2)$$

The rationale behind this definition is that a high probability in the null model for the angular separation between nodes  $i$  and  $j$  to be larger than observed indicates that they are closer than expected in similarity space, hence signaling a highly hierarchical connection. Being a probability,  $h_{ij}$  is bounded in the interval  $[0, 1]$ , while the angular separation between nodes  $\Delta\theta_{ij}^{\text{obs}} \in [0, \pi]$ . Furthermore, Eq. (2) can be computed analytically giving an expression depending on the coordinates of the nodes in the embedding (see Methods D). In particular, in

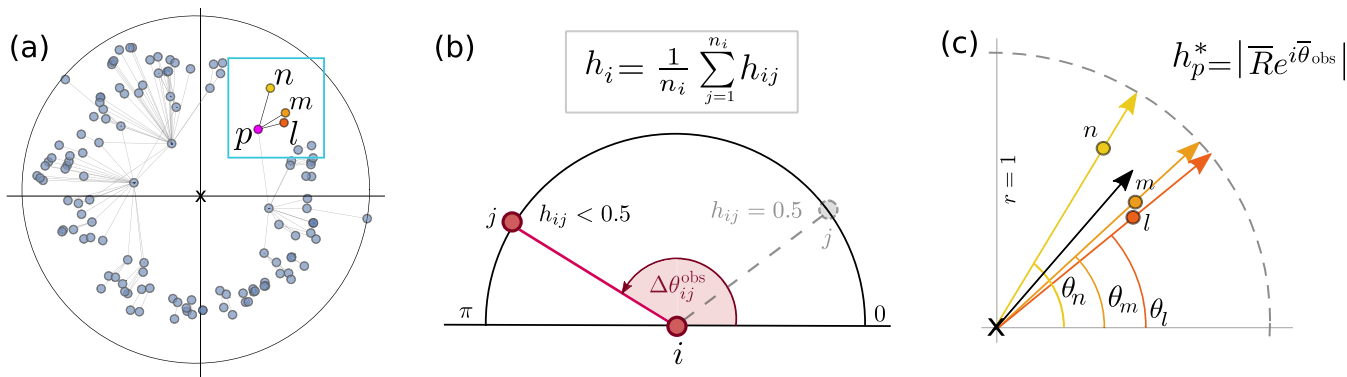


FIG. 1. Hierarchy load in geometric networks. (a) Illustration of a geometric network in the hyperbolic disk. Node  $p$  and its subordinates  $l, m, n$  are highlighted inside a blue square. Notice node  $p$  has a fourth neighbor which is not a subordinate since it lies at a smaller radial position from the center of the disk and thus has higher degree than node  $p$ . (b) Hierarchy load of a link,  $h_{ij}$ , between a generic node  $i$  and a lower-status neighbor  $j$  located at angular distance  $\Delta\theta_{ij}^{\text{obs}}$ . Notice in this example  $h_{ij}$  is necessarily  $h_{ij} < 0.5$  as  $j$  is located at an angular distance greater than the expected from the null model, which appears depicted in grey. The hierarchy load of node  $i$ ,  $h_i$ , is obtained from averaging all link hierarchy loads corresponding to this node. (c) Hierarchy load of node  $p$  from pannel A as the circular mean of the angular coordinates of lower-status neighbors  $l, m, n$ . Mean resultant vector is depicted by a black arrow of  $|\bar{R}| < 1$ . Vectors added in the calculation of  $h_p^*$  [see Eq. (2)] appear as arrows in yellow, orange and red inside a circle of unit radius.

a synthetic network generated with the  $\mathcal{S}^1$  model, the expected value of  $h_{ij}$  for any  $\{i, j\}$  is  $h_{ij} = 1/2$  since, in that case, the observed angular distance  $\Delta\theta_{ij}^{\text{obs}}$  is generated by the model and, hence, it is a random variable distributed according to the same distribution as  $\Delta\theta_{ij}$ . In other words, Eq. (2) reduces to the probability for two equally distributed variables  $a$  and  $b$  to fulfill  $a > b$ . As a consequence, Eq. (2) has the clear advantage of being size independent, in the sense that the value  $h_{ij} = 1/2$  defines the  $\mathcal{S}^1$ -model hierarchy baseline for any  $N$ , therefore allowing us to compare the hierarchy structure of networks of different size.

Link hierarchy loads inform about how substantial is the contribution of a link towards the hierarchy by comparison with the reference level provided by the  $\mathcal{S}^1$  null model, which is  $h_{ij} = 1/2$  as explained above. Accordingly, when the link hierarchy load is higher than the reference,  $h_{ij} > 1/2$ , nodes  $i$  and  $j$  are closer than expected in angular distance, and so they are more similar and their relationship is more hierarchical. In contrast, when the probability is low,  $h_{ij} < 1/2$ ,  $i$  and  $j$  are more dissimilar than expected by the null model, meaning that they lie farther away in the angular space and, thus, their relationship is less hierarchical. In fact,  $h_{ij} = 0$  if  $\Delta\theta_{ij}^{\text{obs}} = \pi$ , while  $h_{ij} = 1$  if  $\Delta\theta_{ij}^{\text{obs}} = 0$ .

Finally, within the same framework we also define a measure of hierarchy load for nodes as

$$h_i = \frac{1}{n_i} \sum_{j=1}^{n_i} h_{ij}, \quad (3)$$

where the sum runs over the  $n_i$  lower-status neighbours  $j$  of node  $i$  satisfying  $\kappa_j < \kappa_i$ .

### B. Hierarchy load of nodes in terms of angular concentration

Alternatively, one can also measure the hierarchy load of a node  $i$  as the angular concentration of its  $n_i$  lower-status neighbors by computing the *circular mean* of their angles  $\{\theta_j\}_{j=1, \dots, n_i}$ . This method is different but very similar to the

node hierarchy measure used in [35] for the analysis of the World Trade Web. When lower-status neighbors are concentrated in a narrow angular sector it means that node  $i$  has a tendency to establish links with more similar nodes, hence node  $i$  is contributing towards a more hierarchical structure and thus carrying a higher hierarchy load  $h_i$ . Conversely, when the lower-status neighbors of  $i$  are distributed in a very broad angular sector, the hierarchy load of node  $i$  is low, indicating that it is able to establish links with more dissimilar lower-status nodes, thus supporting a flatter organization.

The hierarchy load of a node in terms of angular concentration of its  $n$  lower-status neighbors is computed as the length of the mean resultant vector,

$$h^* = \left| \frac{1}{n} \sum_{j=1}^n e^{i\theta_j^{\text{obs}}} \right| = |\bar{R} e^{i\bar{\theta}^{\text{obs}}}| = \bar{R}, \quad (4)$$

see Fig. 1(c). The modulus of the circular mean vector  $\bar{R} \in [0, 1]$  is a good proxy of angular confinement since it is 0 for angles pointing in opposite directions and it becomes 1 when the angles are totally aligned. The measure is simple enough so that it generalizes well to networks of very different domains as long as they admit a geometric interpretation. It is worth mentioning that, since the average angular separation between nodes decreases with the network size  $N$ , this quantity increases with network size. This should be taken into account when comparing  $h^*$  measurements among networks of different size.

### C. Hierarchy load vs geometric communities

To understand how the global distribution of angles, and, in particular, the existence of geometric communities, could affect the spectrum of hierarchy loads, we consider synthetic networks with controllable angular concentration of nodes. Typically, real network maps present angular regions more densely populated than others, which define a partition of the

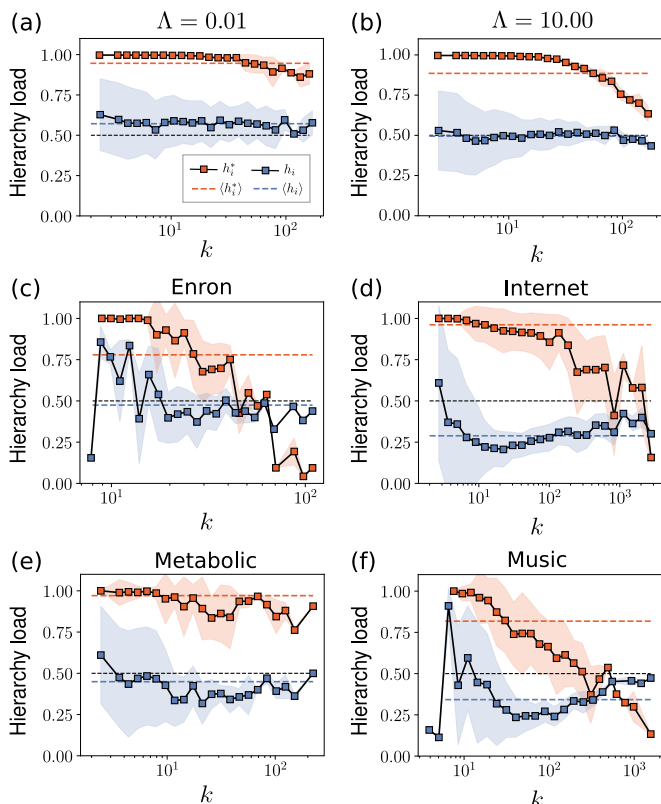


FIG. 2. Hierarchy load spectra of synthetic and real networks. [(a) and (b)] Hierarchy load spectrums for synthetic SCSS networks of size  $N = 1000$  nodes, generated with power-law degree distribution exponent  $\gamma = 2.50$ , clustering parameter  $\beta = 2.50$  and variable attractiveness  $\Lambda$ . Results are averaged over 10 network realizations for each choice of  $\Lambda$ . [(c)–(f)] Hierarchy load spectra of 4 real networks. In all plots blue curves correspond to  $h_i$  in Eq. (3) computed from the link hierarchy loads while red curves correspond to  $h_i^*$  in Eq. (4) as given by the circular mean resultant vector. Dashed blue and dashed red lines indicate the corresponding average hierarchy load values for the whole network. Dashed black lines provide the reference level of the model, placed at hierarchy load  $1/2$ .

network into different geometric groups [20,34–36]. Geometric communities can be accurately detected using algorithms such as the Topological critical gap method [35] and reproduced by geometric network models [37,38]. We generated synthetic networks with tunable geometric communities using the soft communities in similarity space (SCSS) model [37] (see Methods B). The SCSS model is an extension of the  $S^1$  model that enables to create scale-free networks with high clustering while controlling for the global heterogeneity of the distribution of angular coordinates. Essentially, this model relies on a preferential attachment process in similarity space [38], so that the angular coordinates of nodes depend on the angular coordinates of higher-degree nodes. The effect of this similarity preferential attachment is regulated by a parameter  $\Lambda$ . The SCSS recovers the  $S^1$  model in the limit of homogeneous angular distributions, which corresponds to  $\Lambda \rightarrow \infty$ .

We show the spectra of hierarchy loads  $h_i$ , Eq. (3), and  $h_i^*$ , Eq. (4), in Figs. 2(a) and 2(b) for synthetic networks with very different geometric community strengths ( $\Lambda = 0.01$

and  $\Lambda = 10.0$ ). The spectra of hierarchy loads is measured by averaging the node hierarchy loads over degree classes. The two measures display different results, but in both cases, the global angular heterogeneity has a minor effect in shaping the hierarchy loads of nodes. Therefore the spectrum of hierarchy loads of nodes is not merely a measurement of geometric community structure. As expected,  $h_i$  spectra are flat and lie around the average hierarchy load  $\langle h \rangle = \sum_{i=1}^N h_i/N$  for the whole network. At large  $\Lambda$  values,  $\langle h \rangle$  tends to 0.5 because, by construction, SCSS networks recover the  $S^1$  model in this limit. Heterogeneous angular distributions in the limit of small  $\Lambda$  values reduce the average angular distance between nodes in the network, and as a consequence the average hierarchy load of the network is slightly above 0.5.

This effect is also evident in the spectra of hierarchy loads  $h_i^*$ , where  $\langle h \rangle^*$  is lower for more homogeneous angular distributions ( $\Lambda = 10.00$ ,  $\langle h \rangle^* = 0.89 \pm 0.10$ ), as compared with networks with more heterogeneous distributions that present higher values ( $\Lambda = 0.01$ ,  $\langle h \rangle^* = 0.95 \pm 0.06$ ). In this case, the average hierarchy level of the synthetic networks is rather high ( $\langle h \rangle^* \approx 0.9$ ), basically due to sustained large values of hierarchy load for a wide range of node degrees.

#### D. Hierarchy spectrum of real networks

We measured and compared the hierarchy spectrum of several real networks from different domains: the email communication network within the Enron company (Enron) [39], the Internet at the autonomous system level (Internet) [20,40], the one-mode projection onto metabolites of the human metabolic network at the cell level (Metabolic) [34] and the network of chord transitions in western popular music (Music) [41]. For more information see Table I and Methods A.

Real networks show a variety of profiles, see blue curves in Figs. 2(c)–2(f). They also present in all cases an average hierarchy load  $\langle h \rangle$  below the reference of  $1/2$ . In particular, Enron shows the highest average hierarchy load at the node level,  $\langle h \rangle = 0.47 \pm 0.19$ , followed by Metabolic, Music and, lastly the Internet with  $\langle h \rangle = 0.29 \pm 0.20$ . Variations in  $\langle h \rangle$  across networks conform to their distinct spectra. For instance, whereas Music and Internet networks show noticeable fluctuations in  $h_i$  across degree classes, these are milder for Enron and Metabolic networks. In general, however, all networks show a tendency for the lowest degree nodes to have the highest hierarchy loads and for the highest degree nodes, or hubs, to approach  $h_i = 1/2$ . This last observation may be attributable to great heterogeneity in the hierarchy load of links  $h_{ij}$  and the fact that hubs present numerous connections with lower status nodes which are averaged when computing  $h_i$ .

Regarding the hierarchy load of nodes in terms of angular concentration, the Internet and Metabolic are significantly more hierarchical ( $\langle h \rangle^* \approx 1$ ) than Music and Enron. The particular trend followed by each profile has its roots in each network's specific degree sequence, whereas angular communities show again minor influence as revealed by the resemblance between the spectra of the four real networks with the spectra of their corresponding geometrically randomized counterparts (see Methods B and Fig. 6). The randomized counterparts consist in replicas of the real networks where

TABLE I. Properties of the data sets under consideration:  $N$ , size of the network;  $E$  number of edges; parameter  $\beta$  estimated from the embedding of the real network;  $k_{\max}$ , highest degree;  $\langle k \rangle$ , average degree; and  $\langle c \rangle$ , average clustering coefficient;  $\langle h \rangle$ , average hierarchy load;  $\langle h \rangle^*$ , alternative average hierarchy load in terms of angular concentration.

Data set	$N$	$E$	$\beta$	$k_{\max}$	$\langle k \rangle$	$\langle c \rangle$	$\langle h \rangle$	$\langle h \rangle^*$
Enron	182	2097	1.99	109	23.04	0.50	$0.47 \pm 0.19$	$0.77 \pm 0.25$
Internet	23752	58416	1.91	2778	4.92	0.61	$0.29 \pm 0.20$	$0.96 \pm 0.10$
Metabolic	1436	4718	2.15	224	6.57	0.54	$0.45 \pm 0.28$	$0.97 \pm 0.09$
Music	2476	20624	2.30	1566	16.66	0.82	$0.34 \pm 0.25$	$0.81 \pm 0.23$

the distribution of angular coordinates is homogenized, thus eliminating geometric community structure, while the rest of properties are preserved and the replica network remains maximally geometric. Moreover, we note that the hierarchy loads of nodes tend to show strong heterogeneity and an inverse correlation with the degree, so that more popular nodes (higher  $k$ ) contribute more to dilute the hierarchical structure by connecting to less affine lower-popularity neighbors. This trend is remarkably pronounced for the Internet, Enron, and Music, while heterogeneity is less pronounced in Metabolic. In fact, this happens because the Metabolic network presents more modular hubs, less prone to connect with dissimilar nodes due to an unusually marked partition of the similarity space.

### III. THE SIMILARITY FILTER AND HIERARCHICAL BACKBONES OF REAL NETWORKS

Strong heterogeneity is found in real networks if we analyze the contribution of lower-popularity neighbors to the hierarchy load of nodes, see Fig. 3. The link hierarchy load contributions involving low popularity neighbors are represented in Fig. 3 by dots color-coded to the blue end of the scale. We find those dots covering almost all the area of the plots, thus indicating that these link hierarchy loads span the entire range of  $h_{ij}$ , both when the link is shared with another node of low degree or with a node of high degree (right end of the  $x$  axis). This observation reveals that the vast majority of link hierarchy loads in the network are distributed rather heterogeneously. We can take advantage of this heterogeneity to filter out the connections that dominate the hierarchical organization of the network in terms of popularity and similarity, what we name the hierarchical similarity backbone (HSB). A hierarchical similarity backbone contains the links that represent statistically significant contributions with respect to the null hypothesis given by the  $S^1$  model, that is the only model able to produce maximum entropy ensembles constrained by power-law degree distributions and clustering and without nonstructural degree correlations [23].

The link hierarchy load in Eq. (2) measures the probability under the null hypothesis that the similarity distance between a node and a lower-status neighbor is larger than the observed in the embedding of the network, what is known as  $p$  value in statistical inference. By imposing a significance level  $\alpha$ , the links that carry a hierarchy load that are not compatible with the random angular distribution of angles in the  $S^1$  model, and reject the null hypothesis, can be filtered out. A hierarchical similarity backbone is then obtained by preserving all the links that satisfy the criterion  $h_{ij} \geq \alpha$ , while discounting the rest.

As we increase the significance level  $\alpha \in [0, 1]$ , the filter progressively focuses on more relevant links to obtain a sequence of nested subgraphs, each with a more strict condition for a link to belong to the HSB of the network, see an illustration in Fig. 4(a). Noteworthy, since the similarity filter is applied to the links, nodes of any degree may find a place in a very hierarchical backbone if they are found to have significantly strong hierarchical connections.

We tested the performance of the similarity filter by exploring the hierarchical similarity backbones of the four real networks considered in this work. Figure 4(b) shows that, for all real networks, low values of  $\alpha$  reduce the number of links drastically while most of the nodes are preserved in the backbone. Notice that  $\alpha$  increases from right to left in

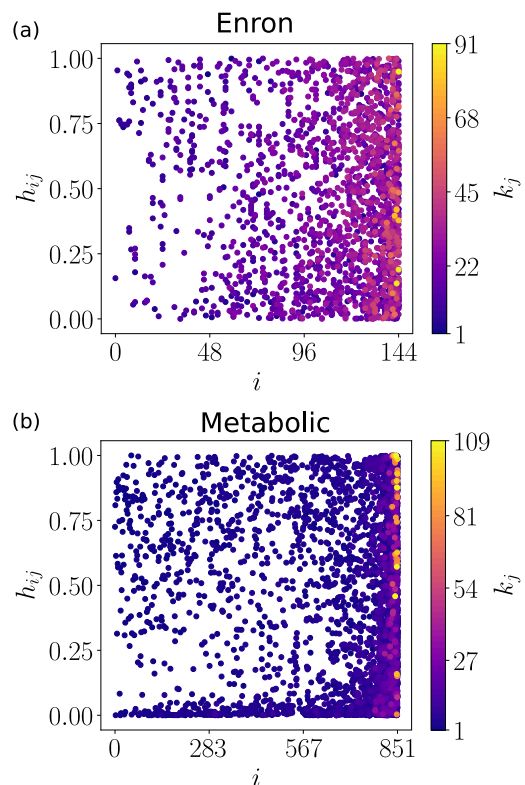


FIG. 3. Link hierarchy loads of real networks. [(a) and (b)] Link hierarchy loads of the Enron and Metabolic networks, respectively. Each dot indicates the value of  $h_{ij}$  of a link established between node  $i$  and one of its lower-status neighbors  $j$  with hidden degree  $\kappa_j$  indicated by the color code. Notice node labels in the  $x$  axis are sorted by increasing hidden degree  $\kappa_i$ , so that more popular  $i$  nodes appear to the right.

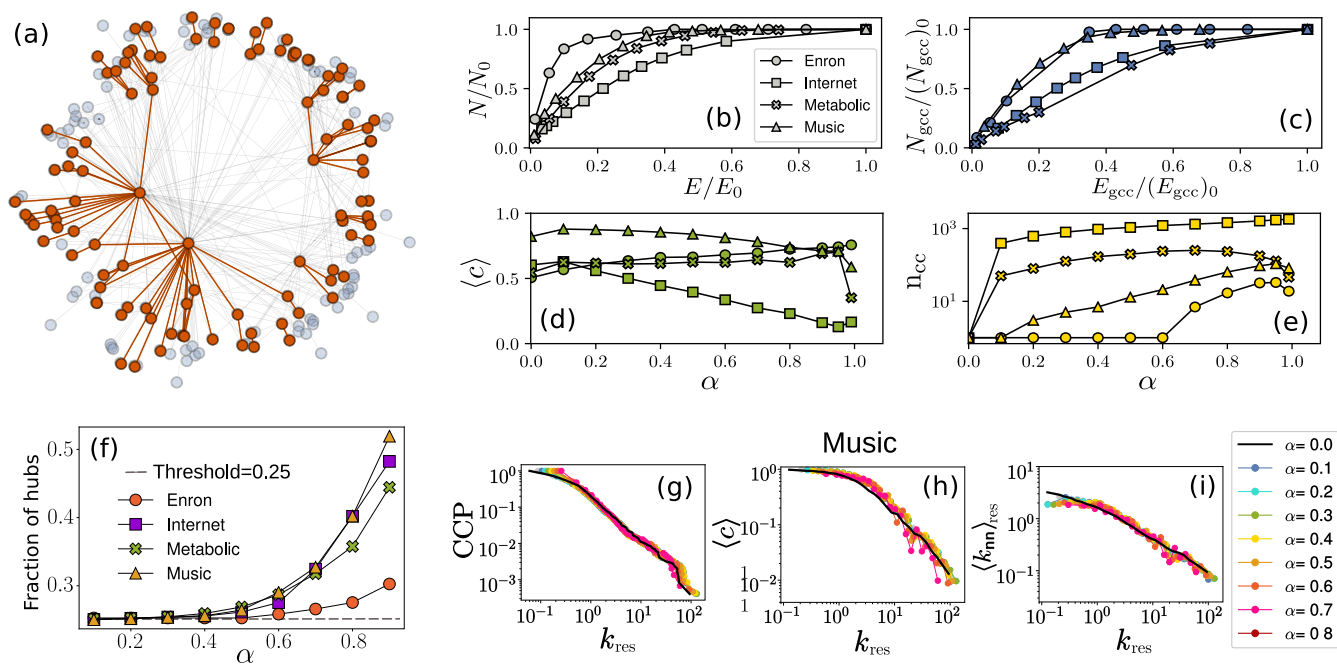


FIG. 4. Hierarchical similarity backbones. (a) A hierarchical similarity backbone of the World Trade Web [35] (in red) filtered with  $\alpha = 0.4$ , on top of the complete network (pale blue). [(b)–(e)] Plots show in order: the relative number of nodes in the backbone, against the relative number of edges in the backbone (subindex 0 refers to the complete network); the relative number of nodes in the backbone giant connected component ( $gcc$ ), against the relative number of edges in the backbone  $gcc$ ; mean clustering coefficient of the backbone for increasing  $\alpha$ ; number of disconnected components against increasing values of the significance  $\alpha$  (f) Fraction of nodes considered hubs, for threshold value  $\tau = 0.25$ , found in backbones obtained with increasing  $\alpha$ . [(g)–(i)] Topological features of the HSBs of the Music network, obtained for  $\alpha$ 's from 0.1 to 0.8. Value  $\alpha = 0.0$  corresponds to the original unaltered network, whereas in the most restrictive HSB ( $\alpha = 0.8$ ) 0.60% of the nodes and 14% of the links remain. From left to right: complementary cumulative probability distribution (CCP) of rescaled degrees,  $k_{res} = k/(k)$ , degree dependent clustering coefficient  $\langle c \rangle$  over rescaled-degree classes, normalized average nearest-neighbor degree  $\langle k_{nn} \rangle_{res} = \langle k_{nn}(k_{res}) \rangle \langle k \rangle / (k^2)$ .

Figs. 4(b) and 4(c). For instance, when filtered with  $\alpha = 0.25$ , the Internet, Metabolic and Music HSBs contain a proportion of edges that is already less than half of the original, so  $E/E_0 < 0.5$ . In contrast, the proportion of nodes in the same backbones remains very high, the lowest case being the Internet, but with still 85% of the nodes. The results in Fig. 4(c) show similar behaviour for the reduction in nodes and edges of the giant connected components ( $gcc$ 's) of the backbones, for the 4 networks under study. Only the decay in number of nodes in the  $gcc$  of the backbones tends to be less abrupt than in Fig. 4(b) and start sooner, at smaller values of the filtering parameter  $\alpha$ . Moreover, we observe in Fig. 4(d) that the mean clustering coefficient of the filtered similarity backbones does not have strong fluctuations and varies little with  $\alpha$ , with the exception of the Internet which shows a clear decreasing trend.

In Fig. 4(f), we inspect the participation of hubs in the HSBs. For this purpose, we sort the nodes in the network from highest degree to lowest and tag as “hubs” all nodes lying within a top slice of the list, delimited by a threshold value  $\tau$ . For instance, when  $\tau = 0.25$ , the top 25% of the nodes in the ranked list are considered as hubs. Subsequently, we keep track of the proportion of such high-degree nodes in every hierarchical backbone for increasing  $\alpha$ . Figure 4(f) demonstrates that, even when considering that a large fraction of the network ( $\tau = 0.25$ ) are nodes of *high* degree, in

fact these nodes only represent, at best, half of the backbone composition [see Internet, Metabolic and Music in Fig. 4(f) for  $\alpha = 0.9$ ]. In Appendix (Fig. 9), we show analogous plots to Fig. 4(f) for a wider range of threshold values  $\tau \in [0.05 - 0.30]$  providing further evidence that, while more restrictive HSBs become enriched with hubs, the similarity backbones still present an assorted composition in terms of node degrees.

Finally, we find that topological features —degree distribution, clustering coefficient and average nearest-neighbor degree— of the original network are preserved in the subgraphs as we increase  $\alpha$  and the backbone is progressively restricted to exceedingly hierarchical links, see Figs. 4(g)–4(i) and 7. Only for very high values of the significance level,  $\alpha \gtrsim 0.8$ , when not only the number of links but also the number of nodes is strongly reduced, the measured topological properties start to deviate from the original curves. This suggests some grade of self-similarity across the sequence of HSBs.

#### IV. GAME DYNAMICS IN HSBs

As an illustration of the importance of HSBs, we study a dynamical process with intriguing behavior in networked systems: the evolution of cooperation. The evolution of cooperation has been studied in different fields [42–45], but it is not well understood yet in scale-free networks [46–50] or in real networks [51], which additionally present high levels

TABLE II. Backbones used to study evolutionary game dynamics. HSB stands for hierarchical similarity backbone, meaning these backbones are obtained using the similarity filter with specific  $\alpha$ . RLR stands for random link removal, thus this method gives random surrogates where a specific number of links have been removed, indicated by LR (links removed). The results in Figs. 5(i)–5(l) and Appendix Fig. 10 are the average of 10 random surrogate realizations. Because of this, and since by construction we fix the number of edges in the random surrogates, only their number of nodes can fluctuate and hence display an error interval.

Network	Filtering method	Threshold	$N_{gcc}$	$E_{gcc}$
Enron	none	–	182	2097
Enron	HSB	$\alpha^o=0.66$	159	587
Enron	HSB	$\alpha^p=0.68$	136	498
Enron	RLR	$LR^o=1510$	$175.30 \pm 1.42$	587
Enron	RLR	$LR^p=1599$	$174.30 \pm 2.45$	498
Internet	none	–	23748	58414
Internet	HSB	$\alpha^o=0.18$	18536	27374
Internet	HSB	$\alpha^p=0.27$	16691	22970
Internet	RLR	$LR^o=31040$	$15901.20 \pm 36.06$	27374
Internet	RLR	$LR^p=35444$	$14172.80 \pm 40.27$	22970
Metabolic	none	–	1436	4718
Metabolic	HSB	$\alpha^o=0.21$	1157	2717
Metabolic	HSB	$\alpha^p=0.25$	1069	2470
Metabolic	RLR	$LR^o=2001$	$1195.60 \pm 14.21$	2717
Metabolic	RLR	$LR^p=2248$	$1147.10 \pm 13.35$	2470
Music	none	–	2476	20624
Music	HSB	$\alpha^o=0.53$	2252	6617
Music	HSB	$\alpha^p=0.59$	2111	5774
Music	RLR	$LR^o=14007$	$1905.60 \pm 12.04$	6617
Music	RLR	$LR^p=14850$	$1824.60 \pm 21.49$	5774

of clustering and finite size effects. In fact, only recently some mechanisms have been proposed to aid cooperation in real networks [52], based precisely in the geometric approach followed in this contribution. Here, we show that, counterintuitively, our HSBs capture the links of real networks that better support cooperative behavior in evolutionary dynamics. We consider an evolutionary prisoner’s dilemma game consisting of two players deciding to cooperate or defect with one another, and gaining a specific reward depending on which of the four possible outcomes takes place. The game proceeds in successive rounds; in each round every node accumulates a payoff  $\pi_i$  resulting from playing the game with all its neighbors; after that, and before moving to the next round, the strategy (cooperate or defect) played by every node is updated simultaneously taking into account an imitation mechanism. The imitation step consist of each node  $i$  deciding to adopt the strategy of a randomly chosen neighbor  $j$  with a probability dependant upon the difference of their payoffs ( $\pi_i - \pi_j$ ), to reflect the tendency of copying more successful neighbors. See Methods C for more details.

We simulated the dynamics on the four real systems analyzed in this work and on two different HSBs for each of them (with  $\alpha$  values and corresponding sizes in number of nodes and links reported in Table II). The results are provided in Figs. 5(a)–5(h). Notice that the similarity backbones are always selected so that they face a considerable reduction in the

number of links while their number of nodes does not decay drastically. That is, for a given real network, the HSBs where we run the game dynamics lie along the slope change part of the blue curves in Figs. 5(i)–5(l), and are identified by blue symbols. We use random surrogates to discern whether the results of the dynamics on HSBs are due to their hierarchical nature. The surrogate backbones are obtained by removing a number of links at random so that they have exactly the same number of links as the corresponding HSBs [see matching fraction of edges between red and blue symbols in Figs. 5(i)–5(l)]. As a result, the fraction of remaining nodes in a given HSB may be higher or lower than in the analogous random surrogate. Regardless of the situation, however, the dynamics results stay consistent, see red curves in Figs. 5(i)–5(l). Note that the fraction of nodes in the random surrogates is still very high even when the number of links has been greatly reduced, akin to the case of similarity backbones. This was expected from the reported robustness of scale-free networks to random removals [53,54].

The evolutionary dynamics are initiated by distributing a proportion of initial cooperators uniformly at random among the nodes of a network. For each of the three graphs (original network and the two similarity backbones), we vary the proportion of initial cooperators and quantify the level of cooperation achieved in the network at the end of the dynamics by measuring the fraction of final cooperators  $N_{coop}^{fin}/N_{gcc}$  after  $10^5$  rounds of the game. Notice no node alters its strategy while playing with its neighbors during an individual round (see Methods C). The fraction of final cooperators  $N_{coop}^{fin}/N_{gcc}$  is averaged over 100 realizations in all showcased curves in Fig. 5 showing the results for the dynamics on HSBs. Notice the system can reach a quenched state before the maximum number of rounds is achieved if all agents become either cooperators or defectors. In that situation, the imitation mechanism does not induce any further evolution and the dynamics becomes effectively frozen.

The results in Figs. 5(a)–5(h) show that the real networks have a tendency for their hierarchical similarity backbones to display final cooperation levels equal or greater than the achieved in the original network for equal proportions of initial cooperators, despite their radically reduced number of links. We expect the dynamics not to work as well as they would in moving the system towards full cooperation when the number of agents to convince is still large but the nodes play the game within small groups of neighbors, as it happens in the backbones. This is because the probability of adopting a neighbor’s strategy,  $P_{i \rightarrow j}$ , drives the evolution towards consensus only when it reflects the tendency to copy a more successful neighbor. This mechanism may be compromised by large fluctuations in the difference of nodes’ collected payoffs ( $\pi_i - \pi_j$ ), when these payoffs come from just few interactions with a small number of neighbors. However, the enhanced final cooperation displayed by the HSBs is observed in general for all networks in Figs. 5(a)–5(h), and specially for Metabolic and Music whose HSBs curves are visibly above the curves for the original networks (in black) for a wide range of initial conditions,  $N_{coop}^{ini}/N_{gcc} \in (0.4-0.9)$ . For instance, the HSB with  $\alpha^p = 0.59$  of the Music network [see blue symbols curve in Fig. 5(h)] has 73% less links than the original network while preserving 83% of nodes and still sustains up to  $\approx 8$

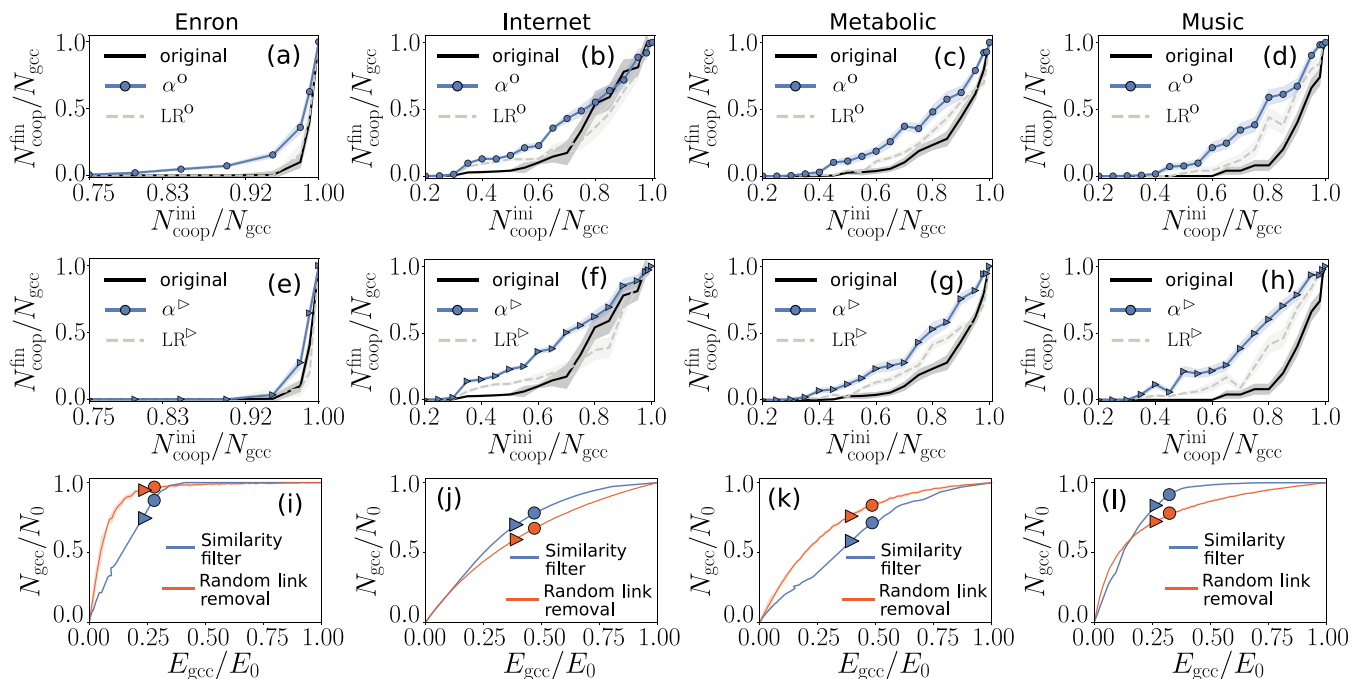


FIG. 5. Evolutionary game dynamics in HSBs of real networks. [(a)–(h)] Fraction of final cooperators against fraction of initial cooperators for similarity backbones of Enron, Internet, Music, and Metabolic networks. For every network we show two plots, each with the results of an HSB filtered with either  $\alpha^o$  or  $\alpha^\triangleright$  (see numeric values in Table II). For an HSB (blue curve with symbols), the corresponding random surrogate appears as a grey dashed line together with the original network in black line for reference. [(i)–(l)] Relative number of nodes against relative number of edges in the *gcc* of similarity backbones, and their corresponding surrogates, for the four networks under study. Blue lines correspond to backbones obtained using the similarity filter and red lines to surrogates obtained by random link removal. Blue circle and triangle symbols highlight the fraction of nodes and edges of the two similarity backbones filtered with  $\alpha^o$  and  $\alpha^\triangleright$ , respectively. The same information is featured by red symbols for the random surrogates.

times more final cooperation than the original network, for a fraction of initial cooperators of 0.8. To further ensure that the enhanced cooperation actually stems from the categorical structure of the backbones one should compare an HSB curve with that of its corresponding random surrogate. By doing so, we observe that the random surrogates happen to reproduce closer the cooperative behaviour of the original network, that is *LR* curves in Figs. 5(a)–5(h) follow the profile of the original network instead of appreciably deviating upwards, thus revealing that the surrogates do not provide a substantial gain in cooperation as opposite to HSBs. In general, the surrogates also require a higher proportion of initial cooperators, around  $N_{\text{coop}}^{\text{ini}}/N_{\text{gcc}} \gtrsim 0.6$ , to produce any sizable increase in final cooperation with respect to the original network. This indicates that the internal hierarchical organization of the HSBs is key to sustain enhanced cooperation. Actually, the similarity filter preferentially removes links with lower hierarchy load, usually consisting of long-range connections established by high degree nodes, whereas the random removal makes no distinction. In fact, given the scale-freeness of real networks, deleting a long-range link at random is less likely due to their scarcity. Therefore, during the dynamics, similarity backbones may develop clusters of same-strategy nodes that are more stable through the evolutionary process than those found in random surrogates, the reason being the former are less exposed to distant contacts belonging to clusters of opposite strategy. This means the hierarchical structure of HSBs enables a better

shielding for the groups of cooperators in the shape of metric clusters [52], which in turn can explain the increased cooperation levels found in similarity backbones.

To additionally validate our results we choose to explore four more different combinations of payoff values for the Music network in Fig. 11. We observe that modifying the pay-offs produces the same qualitative results as discussed above, with HSBs curves clearly surpassing the original network and evidencing that similarity backbones can reach superior cooperation.

### V. DISCUSSION

The existence of a metric space underlying complex networks allows us to provide an enriched interpretation of hierarchy that integrates two dimensions: popularity, or degree rank, and similarity between nodes, thus overcoming the problem of detecting hierarchies in the presence of clustering and the small world effect. The metric approach enables a unified framework to define the hierarchy loads of nodes and links.

Interestingly, the spectra of hierarchy loads of real networks revealed that, in general, these networks are less hierarchical than the reference provided by the maximum entropy null model and show greater variation in the hierarchy load of nodes across degree classes. Particularly, the lowest degree nodes typically contribute more towards the hierarchical structure, although their fluctuations are remarkable.



Moreover, we introduced the *similarity filter*, a link pruning method which exploits the heterogeneity found in the hierarchy load of links. The filter extracts the connections that dominate the hierarchical structure of networks in terms of popularity and similarity, providing what we name hierarchical similarity backbones (HSB). The analysis of such backbones uncovered that, strikingly, the similarity filter is able to preserve network topological features at all scales while discarding a large number of links. Accordingly, from a fundamental point of view, hierarchical backbones could help provide new insight about the percolation properties of highly stratified real networks, aiding control of cascading failures, as well as have the potential to become a standard methodology for the detection, visualization and inspection of hierarchical clusters [55] in machine learning and data science environments.

From a practical point of view, the similarity filter has proven to be an exceptional tool to unravel the backbone that sustains enhanced cooperation in social dilemmas on structured populations. This is in line with previous simulations of prisoners dilemma type dynamics on adaptive networks, showing that cooperation combined with imitation can lead to a hierarchical structure [11]. When this dynamics is played on heterogeneous contact networks with underlying metric structure, the evolution of cooperation leads to the formation of clusters of cooperators in the similarity subspace [52]. In the presence of these clusters, heterogeneity in the degrees was nevertheless found to hinder cooperation. Those findings reveal a tension between the popularity and similarity dimensions in evolutionary dynamics modeling social dilemmas. Our findings here solve this opposition by identifying the similarity backbones composed of significant links that are simultaneously hierarchical in terms of popularity and similarity, and which are expressly relevant in supporting and fostering cooperation.

Lastly, the methods developed in this contribution can be used to study the hierarchical nature of complex networks of any domain as long as they admit a geometric representation. The detection of hierarchical similarity backbones could for instance help in designing controllability of gene regulatory networks, improve communicability in information systems and infrastructures or assess robustness to species loss in ecological networks. Other possibilities for our framework include its extension to multiplex networks, opening promising future lines of research.

## VI. METHODS

### A. Empirical data

All real complex networks used in this paper have been mapped into their hyperbolic latent geometry using the embedding method Mercator [21]. This method mixes machine learning and maximum likelihood approaches to infer the coordinates of the nodes in the underlying hyperbolic disk, while ensuring best congruency between the real network topology and the  $\mathcal{S}^1$  geometric model.

Enron. The network captures the email communication activity (125 409 emails) within employees from the Enron company. Edges are established between email addresses

that shared correspondence. We use the dataset provided in Ref. [56] which includes also information about the organizational roles of 130 users.

Internet. We use the adjacency data for the Internet at the autonomous systems (AS) level assembled by the Archipelago project [40] during June 2009.

Human metabolic. This network is the one-mode projection of metabolites of the bipartite metabolic network of human cell metabolisms. In this representation [34], there is a link between two metabolites if they participate in the same biochemical reaction.

Music. Nodes of the network are chords-sets of musical notes (see Ref. [41]) played in a single beat while edges represent detected transitions between these chords. We use a sparser, undirected version of the network reconstructed in Ref. [33].

## B. Models of network geometry

### 1. $\mathcal{H}^2$ model

An isomorphism exists between the  $\mathcal{S}^1$  and the  $\mathcal{H}^2$  models [22], so that hidden degrees  $\kappa$  are mapped into radial coordinates,  $r$ , in a hyperbolic disk of radius  $R_{\mathcal{H}^2}$ , such that  $\kappa \sim e^{(R_{\mathcal{H}^2} - r)/2}$ . Consequently, in the hyperbolic version, nodes with larger radial coordinates are located towards the edge of the hyperbolic disk and show lower expected degree. Particularly, in the  $\mathcal{H}^2$  model, every node  $i$  is defined by the tuple  $(r_i, \theta_i)$ , and the probability that a link exists between two nodes  $i$  and  $j$  depends on their distance  $d_{ij}$ , as measured in the hyperbolic space using the hyperbolic law of cosines  $\cosh(d_{ij}) = \cosh r_i \cosh r_j - \sinh r_i \sinh r_j \cos \Delta \theta_{ij}$ . Nodes closely positioned in the hyperbolic disk have higher chances of being connected, thus the connection probability  $p(d_{ij})$  must be a decreasing function of distance between them and, specifically, it can be chosen to be

$$p(d_{ij}) = \frac{1}{1 + \exp\left[\frac{\beta(d_{ij} - R_{\mathcal{H}^2})}{2}\right]}, \quad (5)$$

where the parameter  $\beta$  still controls the network's clustering coefficient. In this paper, we mainly use the  $\mathcal{S}^1$  model for calculation purposes, and its equivalent  $\mathcal{H}^2$  version for visualization tasks.

### 2. Soft Communities in similarity space (SCSS)

The SCSS model [37] produces synthetic geometric networks with inhomogeneous angular distributions, derived from geometric preferential attachment mechanisms, which were conceived in growing geometric network models [38,57]. This means, for a network represented in an underlying (hyperbolic) metric space, the initial attractiveness of different angular regions during a geometric preferential attachment process is controlled by a parameter  $\Lambda$ . Consequently, this parameter regulates the heterogeneity of the angular coordinate, so that heterogeneity is a decreasing function of  $\Lambda$ , with  $\Lambda \rightarrow \infty$  recovering the homogeneous distribution. The SCSS model, then takes such heterogeneous angular distribution defined by  $\Lambda$  and adjusts it to an independent power-law degree distribution ( $P(k) \sim k^{-\gamma}$ ) and a tunable level of mean clustering  $\langle c \rangle$ , controlled through parameter  $\beta$ .

The SCSS model does so by introducing correlations between  $\kappa$  and  $\theta$  coordinates of nodes of the geometric network.

### C. Geometric Randomization (GR)

The GR [58] is a model for the randomization of complex networks with geometric structure, which allows to uniformize their angular coordinate distribution, while preserving the exact degree sequence of the network. It thus applies to both real and synthetic networks where nodes have an observed degree and exist in a similarity space. In the GR model, angular coordinates  $\theta$  are assigned to the nodes, chosen uniformly at random from  $[0, 2\pi]$ . The network is then rewired following a likelihood maximization process that ensures the new topology is one generated by the  $S^1$  model, while the observed degrees (and thus the number of edges) remain unaltered. The model is implemented using a single parameter  $\beta$  controlling the mean clustering of the resultant rewired network. The rewiring and maximization procedure executed by the GR are specially useful to produce faithful real network replicas where only geometric (soft) communities have been suppressed.

### D. Evolutionary prisoner’s dilemma

The evolutionary prisoner’s dilemma game [27], conducted on a network, considers that individual nodes playing with their contacts choose to either cooperate (C) or defect (D) every turn. The choice of strategies of the two interacting agents leads to specific payoffs, summarized by the payoff-matrix

$$A = \begin{array}{c|cc} & C & D \\ \hline C & R & S \\ D & T & P \end{array}. \quad (6)$$

That is, if both players cooperate, they both receive the *reward*  $R$  for cooperating. If both players defect, they both receive the *punishment* payoff  $P$ . Lastly, if one of them defects while the other agent cooperates, the defector receives the *temptation* payoff  $T$ , while the cooperator receives the “*sucker’s*” payoff,  $S$ . In order for the game to be recognized as a prisoners’s dilemma the condition  $T > R > P > S$  must apply. Different ordinalities of the parameters define further classes of games [29]. In this paper, the prioner’s dilemma is defined with parameter values:  $T = 1.5$ ,  $R = 1$ ,  $P = 0$ , and  $S = -0.5$ . Further parameter values are explored in Appendix.

The game proceeds in successive rounds. After each round, the strategies (C or D) of all nodes are updated synchronously according to the outcome of the imitation dynamics [27,59,60] that outline the evolutionary mechanism. This means, during a single round, each individual node  $i$  collects payoffs given by Eq. (6) from the interactions with all its neighbors and obtains an accumulated payoff  $\pi_i$ . All players chose then between their old strategy and the strategy of a randomly picked up neighbor  $j$ . In this way, node  $i$  will adopt  $j$ ’s strategy with a probability that depends upon the difference between collected payoffs of both nodes ( $\pi_i - \pi_j$ ) as

$$p_{i \rightarrow j} = \frac{1}{1 + e^{(\pi_i - \pi_j)/\tau}}, \quad (7)$$

which reflects the popular tendency of individuals to copy more successful neighbors. Such updating rule is a common choice in evolutionary dynamics [61], known as the Fermi rule since it is based in the Fermi distribution from statistical mechanics. The variable  $(\frac{1}{\tau})$ , which in Physics stands for the inverse temperature, can be interpreted here as the intensity of the selection. That is, parameter  $(\frac{1}{\tau})$  (which we set to 0.5) controls the noise added to the decision-making process of, otherwise, perfectly rational players. After the simultaneous update of strategy of all nodes, their accumulated payoffs are reset and a new round begins.

### E. Probability distribution of angular distances between connected nodes in the $S^1$ model

The hierarchy load of an observed link,  $h_{ij}$ , can be computed analytically given the angular distance between the two nodes observed in the embedding,  $\Delta\theta_{ij}^{\text{obs}}$ . To derive a closed expression for it, we first need to write down an expression for the probability distribution function of the angular separation between two nodes in the  $S^1$  model *conditioned to the fact that they are connected*. Using Bayes’ rule, we see that

$$p(\Delta\theta_{ij}|a_{ij} = 1) = \frac{p(a_{ij} = 1|\Delta\theta_{ij})p(\Delta\theta_{ij})}{p(a_{ij} = 1)}, \quad (8)$$

where  $a_{ij}$  is the adjacency matrix element corresponding to the two nodes. In the above expression,  $p(a_{ij} = 1|\Delta\theta_{ij})$  is the connection probability, Eq. (1), while the distribution of angular distances is simply  $p(\Delta\theta_{ij}) = 1/\pi$ , given that angular coordinates are homogeneously distributed in the  $S^1$  model. The denominator can be obtained by direct integration,  $p(a_{ij} = 1) = \int_0^\pi p(a_{ij} = 1|\Delta\theta_{ij})p(\Delta\theta_{ij})d\Delta\theta_{ij}$ . Using the definition of  $h_{ij}$ , we obtain

$$\begin{aligned} h_{ij} &= P(\Delta\theta_{ij} > \Delta\theta_{ij}^{\text{obs}}) \\ &= 1 - \int_0^{\Delta\theta_{ij}^{\text{obs}}} p(\Delta\theta_{ij}|a_{ij} = 1)d\Delta\theta_{ij} \\ &= 1 - \frac{\int_0^{\Delta\theta_{ij}^{\text{obs}}} p(a_{ij} = 1|\Delta\theta_{ij})p(\Delta\theta_{ij})d\Delta\theta_{ij}}{\int_0^\pi p(a_{ij} = 1|\Delta\theta_{ij})p(\Delta\theta_{ij})d\Delta\theta_{ij}} \\ &= 1 - \frac{\Delta\theta_{ij}^{\text{obs}} {}_2F_1\left(1, \frac{1}{\beta}; 1 + \frac{1}{\beta}; -\left(\frac{R\Delta\theta_{ij}^{\text{obs}}}{\mu\kappa_i^{\text{obs}}\kappa_j^{\text{obs}}}\right)^\beta\right)}{\pi {}_2F_1\left(1, \frac{1}{\beta}; 1 + \frac{1}{\beta}; -\left(\frac{R\pi}{\mu\kappa_i^{\text{obs}}\kappa_j^{\text{obs}}}\right)^\beta\right)}, \end{aligned} \quad (9)$$

where  ${}_2F_1(a, b; c; z)$  is the hypergeometric function. Equation (9) yields the hierarchy load of a given link in terms of the angular separation and product of hidden degrees of the nodes at both ends of the edge, as well as of the global parameters  $R$ ,  $\mu$ , and  $\beta$ .

Finally, let us also show that the expected value of  $h_{ij}$  for any link in a network generated by the  $S^1$  model is  $\langle h_{ij} \rangle_{S^1} = 1/2$ . To calculate  $\langle h_{ij} \rangle_{S^1}$ , we only need to notice that, in this case, the angular separation between the nodes at the ends of the edge in the resulting network,  $\Delta\theta_{ij}^{\text{obs}}$ , is itself a random variable with distribution  $\rho(\Delta\theta_{ij}^{\text{obs}})$  given by Eq. (8) (that is,  $\rho(\Delta\theta_{ij}^{\text{obs}}) = p(\Delta\theta_{ij}^{\text{obs}}|a_{ij} = 1)$ ) and, therefore, the expected

value of  $h_{ij}$  is

$$\begin{aligned}
 \langle h_{ij} \rangle_{S^1} &= \langle P(\Delta\theta_{ij} > \Delta\theta_{ij}^{\text{obs}}) \rangle_{S^1} \\
 &= \int_0^\pi \rho(\Delta\theta_{ij}^{\text{obs}}) P(\Delta\theta_{ij} > \Delta\theta_{ij}^{\text{obs}}) d\Delta\theta_{ij}^{\text{obs}} \\
 &= \int_0^\pi \rho(\Delta\theta_{ij}^{\text{obs}}) \int_{\Delta\theta_{ij}^{\text{obs}}}^\pi \rho(\Delta\theta_{ij}) d\Delta\theta_{ij} d\Delta\theta_{ij}^{\text{obs}} = \frac{1}{2}.
 \end{aligned} \tag{10}$$

In the last step, we have used the fact that  $\rho(z)$  is normalized,  $\int_0^\pi \rho(z) dz = 1$ .

Codes and data supporting the findings of this study are available from the corresponding author upon request.

### ACKNOWLEDGMENTS

M.Á.S. acknowledge support from a James S McDonnell Foundation Scholar Award in Complex Systems; Ministerio de Ciencia, Innovación y Universidades of Spain Project No. FIS2016-76830-C2-2-P (AEI/FEDER, UE); the project Mapping Big Data Systems: embedding large complex networks in low-dimensional hidden metric spaces—Ayudas Fundación BBVA a Equipos de Investigación Científica 2017, and the Generalitat de Catalunya Grant No. 2017SGR106. G.G.-P. acknowledges financial support from the Academy of Finland via the Centre of Excellence program (Project No. 312058 as well as Project No. 287750), and from the emmy.network foundation under the aegis of the Fondation de Luxembourg.

E.O., G.G.-P., and M.Á.S. contributed to the design and implementation of the research, to the analysis of the results and to the writing of the manuscript.

The authors declare no competing financial interests.

### APPENDIX

#### 1. Hierarchy load spectrums of nodes for GR network replicas

Figures 2(c)–2(f) shows the spectrum of hierarchy loads of nodes in terms of angular concentration,  $h_i^*$ , for 4 real networks. Here, we provide the same metric for angularly randomized versions of such real networks, obtained using the Geometric Randomization model (see Methods B in main paper). By comparing the spectra in Figs. 2(c)–2(f) and 6 in this Appendix, one can notice that the complete homogenization of the angular coordinates of nodes (and thus the full elimination of geometric communities) does not translate into radical changes in the hierarchy load profiles. Instead, below (Fig. 6) we observe for each of the randomized networks that the inverse correlation with the degree of the hierarchy loads is compatible with that of the original network, and that the average hierarchy load  $\langle h \rangle^*$  values of the GR networks are not remarkably different to that of the real ones. The most noticeable influence of geometric community structure in the hierarchy load measurement is found for the Metabolic network, which turns to be a natural effect given the pronounced bimodal distribution of node angles of the original real network.

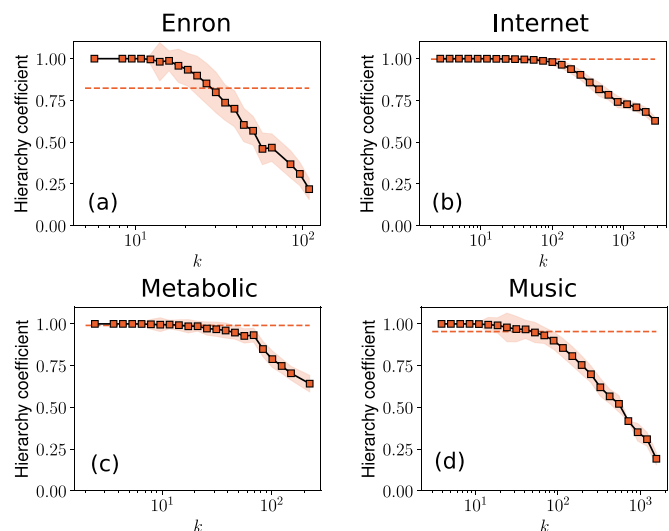


FIG. 6. Hierarchy load spectrums for Geometric randomization (GR) model replicas of each of the 4 real networks under study. Each spectrum is the average over 10 independent realizations. Dashed lines indicate the average hierarchy load of the network, obtained as  $\langle h \rangle^* = N^{-1} \sum_{i=1}^N h_i^*$ , yielding to the following values for each network:  $\langle h \rangle_{\text{Enron}}^* = 0.82 \pm 0.19$ ,  $\langle h \rangle_{\text{Internet}}^* = 0.99 \pm 0.02$ ,  $\langle h \rangle_{\text{Metabolic}}^* = 0.99 \pm 0.04$ , and  $\langle h \rangle_{\text{Music}}^* = 0.95 \pm 0.11$ .

#### 2. Topological properties of similarity backbones of real networks

First, in Fig. 7, we provide results for the topological features of similarity backbones examined in Figs. 4(g)–4(i) for the rest of real networks under study. Secondly, in Fig. 8, we cover further topological metrics of the HSBs and showcase them against the  $\alpha$  parameter controlling the filtering procedure.

#### 3. Composition of similarity backbones of real networks

In this section, we sort the nodes in the network from highest degree to lowest and tag as “hubs” all nodes lying within a top slice of the list, delimited by a threshold value  $\tau$ . Subsequently, in Fig. 9, we keep track of the proportion of such high-degree nodes in every similarity backbone of increasing  $\alpha$  for the 4 real networks. The results are analogous to those of Fig. 4(f) but obtained for an extended range of  $\tau$  values.

#### 4. Evolutionary prisoner’s dilemma game dynamics

First, for the four real networks analysed, we report in Table S1 the number of nodes and edges in the *gccs* of every similarity backbone and random surrogate along with  $\alpha$  filtering values and number of links removed, respectively. Secondly, in Fig. 10, we provide analogous plots of those in Figs. 4(g)–4(i) and 7 for the random surrogates, showcasing their main topological features. Lastly, in Fig. 11, we explore the dynamics on similarity backbones (HSB) for further payoff values for one of the networks (Music).

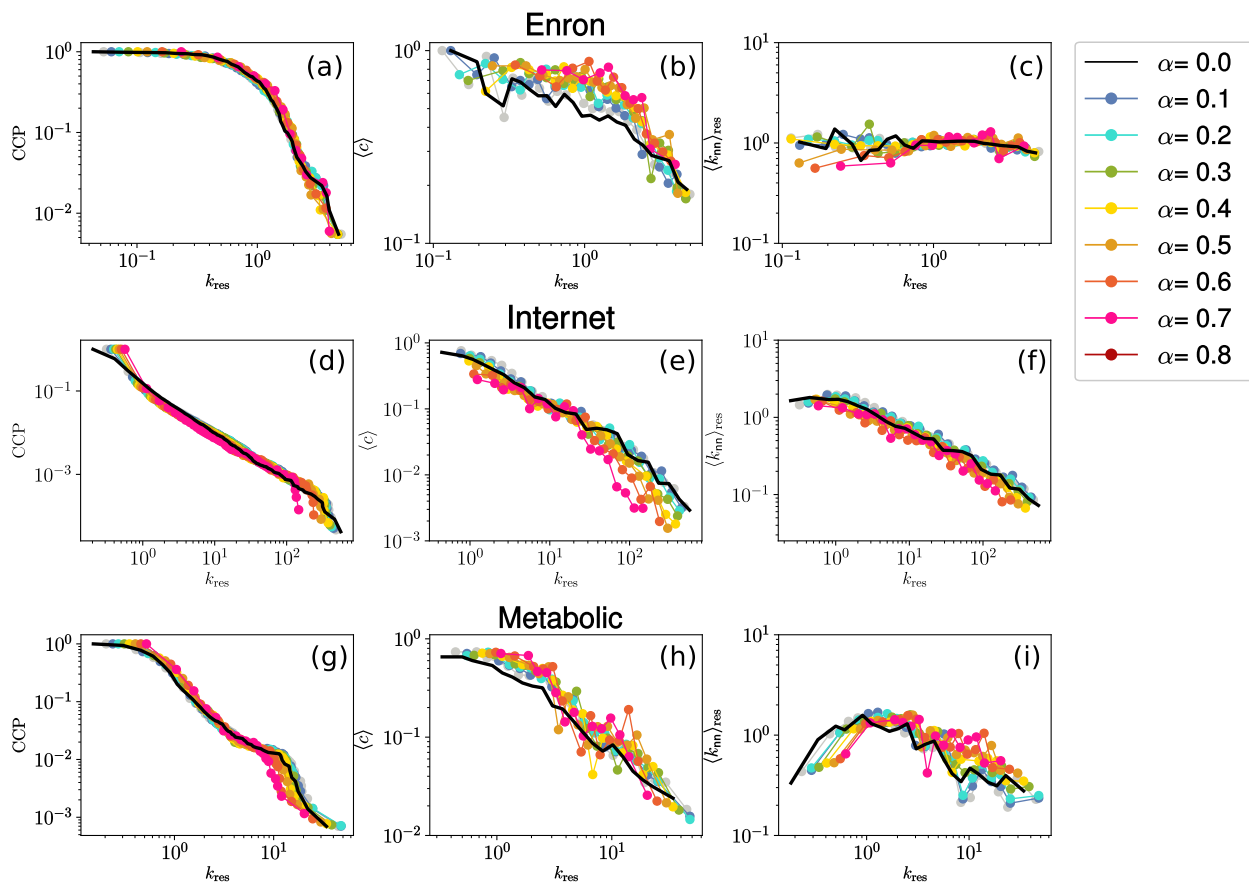


FIG. 7. Topological features of the hierarchical backbones of [(a)–(c)] Enron, [(d)–(f)] Internet, and [(g)–(i)] Metabolic, obtained for  $\alpha \in [0.1–0.8]$ . In each case, value  $\alpha = 0.0$  corresponds to the original network. The properties in each row, from left to right, are complementary cumulative distribution of rescaled degrees ( $k_{res} = k/\langle k \rangle$ ), degree dependent clustering coefficient over rescaled-degree classes, and normalized average nearest-neighbor degree  $\langle k_{nn} \rangle_{res} = \langle k_{nn}(k_{res}) \rangle \langle k \rangle / \langle k^2 \rangle$ .

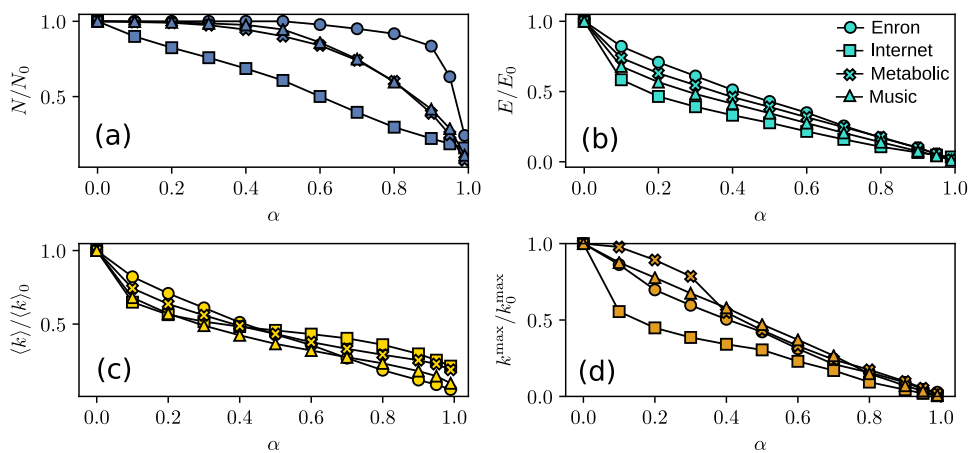


FIG. 8. Topological properties of hierarchical backbones of the four real networks against increasingly restrictive filtering parameter ( $\alpha$ ) values. Plots from (a) to (d) show, for increasing values of  $\alpha$ : the normalized number of nodes in the backbone; the normalized number of edges in the backbone; normalized average degree of the backbone and the HSB normalized maximum degree.

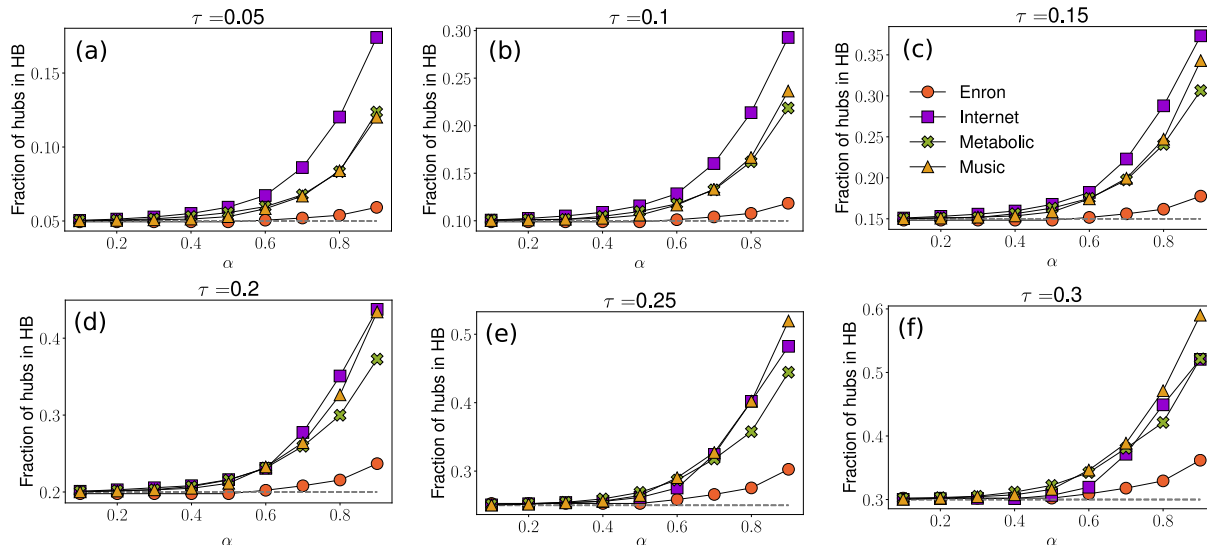


FIG. 9. Fraction of nodes considered hubs for threshold values in range  $\tau = [0.05-0.30]$ , found in backbones obtained with increasing  $\alpha$  for the real networks Enron, Internet, Metabolic, and Music.

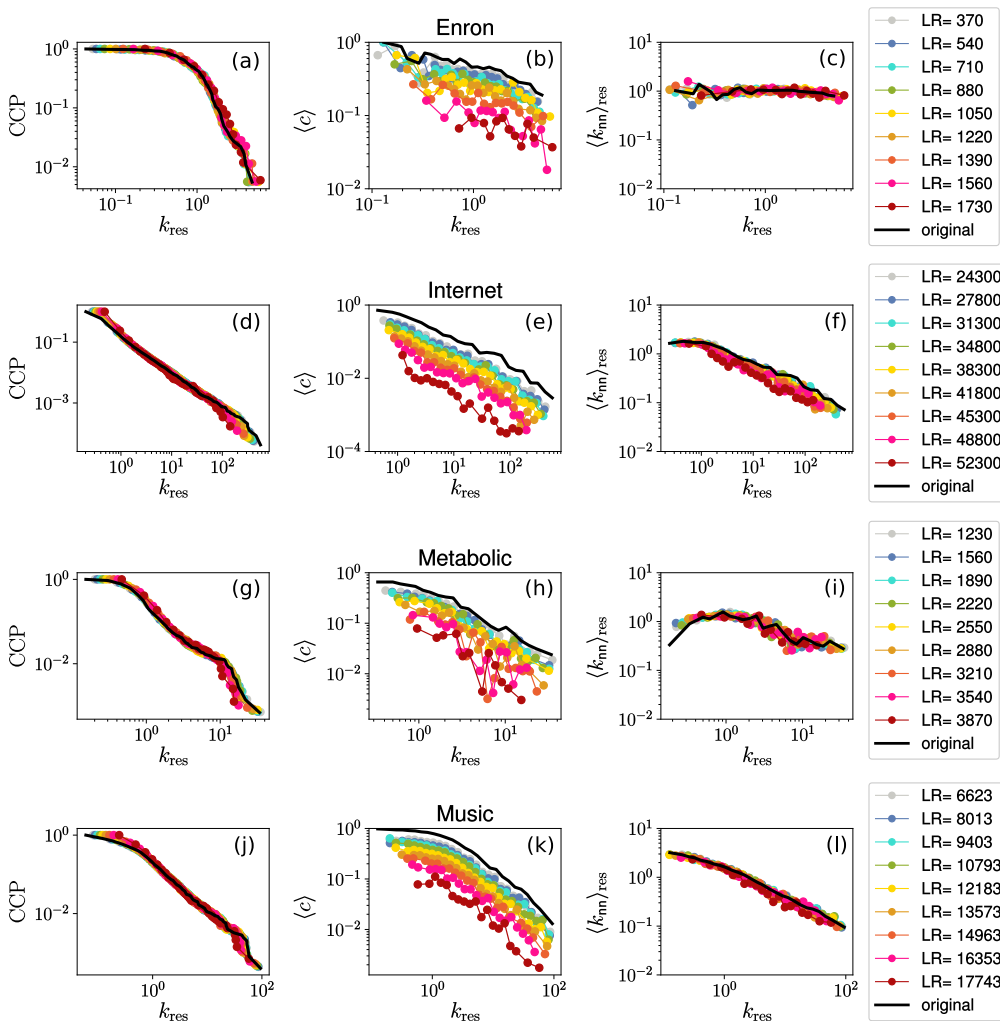


FIG. 10. Topological features of random surrogates and original network of [(a)–(c)] Enron, [(d)–(f)] Internet, [(g)–(i)] Metabolic, and [(j)–(l)] Music. In each case, value  $LR$  indicates the number of links that have been removed at random to produce the surrogate backbone. The properties in each row, from left to right, are: complementary cumulative distribution of rescaled degrees  $k_{res} = k/\langle k \rangle$ , degree dependent clustering coefficient over rescaled-degree classes, normalized average nearest-neighbor degree  $\langle k_{nn} \rangle_{res} = \langle k_{nn}(k_{res}) \rangle \langle k \rangle / \langle k^2 \rangle$ .

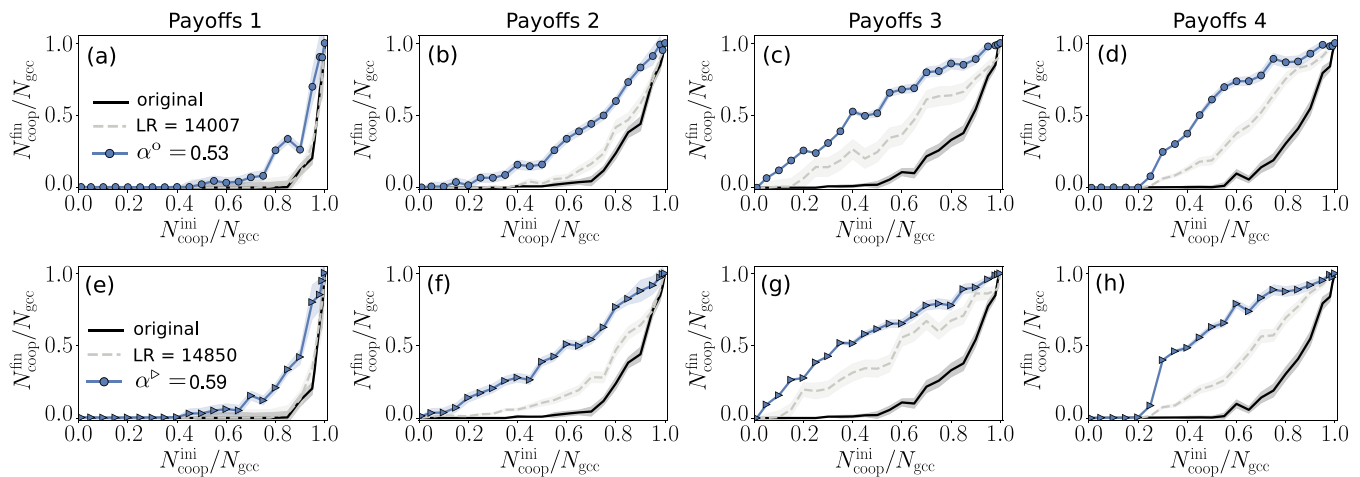


FIG. 11. Fraction of final cooperators against fraction of initial cooperators for similarity backbones and random surrogates of Music network under different payoffs. The results of the dynamics in each plot are obtained using the following payoff values for the evolutionary prisoner’s dilemma game: (a) and (e)  $T = 1.0$ ,  $R = 0.5$ ,  $P = -0.5$ ,  $S = -1.0$  (b) and (f)  $T = 1.5$ ,  $R = 1.0$ ,  $P = 0.5$ ,  $S = 0.0$  (c) and (g)  $T = 2.0$ ,  $R = 1.5$ ,  $P = 0.5$ ,  $S = 0.0$  (d) and (h)  $T = 2.0$ ,  $R = 1.5$ ,  $P = 0.0$ ,  $S = -0.5$ . Results for the original network appear for reference, indicated by a black line. HSB filtered with  $\alpha = 0.53$  are shown as blue curves with circles whereas HSBs filtered with  $\alpha = 0.59$  correspond to blue curves with triangles in the second row. Results of random surrogates are displayed as grey dashed lines.

[1] B. Corominas-Murtra, J. Goñi, R. V. Solé, and C. Rodríguez-Caso, On the origins of hierarchy in complex networks, *Proc. Natl. Acad. Sci. USA* **110**, 13316 (2013).

[2] D. H. Erwin and E. H. Davidson, The evolution of hierarchical gene regulatory networks, *Nat. Rev. Genet.* **10**, 141 (2009).

[3] B. Johnson and J. R. Carey, Hierarchy and connectedness as determinants of health and longevity in social insects, in *Sociality, Hierarchy, Health: Comparative Biodemography: A Collection of Papers* (National Academies Press, Washington DC, 2014), pp. 269–294.

[4] A. Samuels, J. B. Silk, and P. S. Rodman, Changes in the dominance rank and reproductive behavior of male bonnet macaques (*Macaca radiata*), *Anim. Behav.* **32**, 994 (1984).

[5] P. Saunders, *Social Class and Stratification*, Society now (Routledge and Kegan Paul, Abingdom, Oxon, UK, 1990).

[6] M. DeLoura, *Game Programming Gems 2*, Charles River Media programming (Charles River Media, Hingham, Massachusetts, 2001).

[7] D. Allerton, *Essentials of Grammatical Theory: A Consensus View of Syntax and Morphology* (Routledge and Kegan Paul, Abingdom, Oxon, UK, 1979).

[8] M. Yeston, *The Stratification of Musical Rhythm* (Yale University Press, London, UK, 1976).

[9] E. Mones, L. Vicsek, and T. Vicsek, Hierarchy measure for complex networks, *PLoS ONE* **7**, e33799 (2012).

[10] G. Theraulaz, E. Bonabeau, and J.-L. Deneubourg, Self-organization of hierarchies in animal societies: The case of the primitively eusocial wasp *Polistes dominulus*, *J. Theor. Biol.* **174**, 313 (1995).

[11] V. M. Eguíluz, M. G. Zimmermann, C. J. Cella-Conde, and M. S. Miguel, Cooperation and the emergence of role differentiation in the dynamics of social networks, *Am. J. Sociol.* **110**, 977 (2005).

[12] E. Ravasz, and A.-L. Barabási, Hierarchical organization in complex networks, *Phys. Rev. E* **67**, 026112 (2003).

[13] L. da Fontoura Costa, The Hierarchical Backbone of Complex Networks, *Phys. Rev. Lett.* **93**, 098702 (2004).

[14] N. Tatti, Hierarchies in directed networks, in *Proceedings of the 2015 IEEE International Conference on Data Mining, Atlantic City, NJ* (IEEE, Piscataway, NJ, 2015), pp. 991–996.

[15] E. Letizia, P. Barucca, and F. Lillo, Resolution of ranking hierarchies in directed networks, *PLoS ONE* **13**, e0191604 (2018).

[16] N. Tatti, Dynamic hierarchies in temporal directed networks, in *Machine Learning and Knowledge Discovery in Databases: ECML PKDD 2018*, edited by M. Berlingerio, F. Bonchi, T. Gärtner, N. Hurley, and G. Iffrim, Lecture Notes in Computer Science Vol. 11052 (Springer, Cham, 2019), pp. 55–70.

[17] A. Trusina, S. Maslov, P. Minnhagen, and K. Sneppen, Hierarchy Measures in Complex Networks, *Phys. Rev. Lett.* **92**, 178702 (2004).

[18] N. Perra and S. Fortunato, Spectral centrality measures in complex networks, *Phys. Rev. E* **78**, 036107 (2008).

[19] M. Á. Serrano, D. Krioukov, and M. Boguñá, Self-Similarity of Complex Networks and Hidden Metric Spaces, *Phys. Rev. Lett.* **100**, 078701 (2008).

[20] M. Boguñá, F. Papadopoulos, and D. Krioukov, Sustaining the Internet with hyperbolic mapping, *Nat. Commun.* **1**, 62 (2010).

[21] G. García-Pérez, A. Allard, M. Á. Serrano, and M. Boguñá, Mercator: Uncovering faithful hyperbolic embeddings of complex networks, *New J. Phys.* **21**, 123033 (2019).

[22] D. Krioukov, F. Papadopoulos, M. Kitsak, A. Vahdat, and M. Boguñá, Hyperbolic geometry of complex networks, *Phys. Rev. E* **82**, 036106 (2010).

[23] M. Boguñá, D. Krioukov, P. Almagro, and M. Serrano, Small worlds and clustering in spatial networks, *Phys. Rev. Res.* **2**, 023040 (2020).

- [24] M. Á. Serrano, M. Boguñá, and A. Vespignani, Extracting the multiscale backbone of complex weighted networks, *Proc. Natl. Acad. Sci. USA* **106**, 6483 (2009).
- [25] H. A. Simon, The architecture of complexity, *Proc. Am. Philos. Soc.* **106**, 467 (1962).
- [26] P. E. Agre, Hierarchy and history in Simon's "architecture of complexity", *J. Learn. Sci.* **12**, 413 (2003).
- [27] G. Szabó and G. Fáth, Evolutionary games on graphs, *Phys. Rep.* **446**, 97 (2007).
- [28] G. Szabó and C. Tóke, Evolutionary prisoner's dilemma game on a square lattice, *Phys. Rev. E* **58**, 69 (1998).
- [29] M. Nowak, *Evolutionary Dynamics: Exploring the Equations of Life* (Harvard University Press, Cambridge, MA, 2006).
- [30] J. Weibull, *Evolutionary Game Theory* (MIT Press, Cambridge, MA, 1998).
- [31] A. Gulyás, J. J. Bíró, A. Kőrösi, G. Rétvári, and D. Krioukov, Navigable networks as nash equilibria of navigation games, *Nat. Commun.* **6**, 7651 (2015).
- [32] E. Ortiz, M. Starnini, and M. Á. Serrano, Navigability of temporal networks in hyperbolic space, *Sci. Rep.* **7**, 15054 (2017).
- [33] G. García-Pérez, M. Boguñá, and M. Á. Serrano, Multi-scale unfolding of real networks by geometric renormalization, *Nat. Phys.* **14**, 583 (2018).
- [34] M. Á. Serrano, M. Boguna, and F. Sagues, Uncovering the hidden geometry behind metabolic networks, *Mol. BioSyst.* **8**, 843 (2012).
- [35] G. García-Pérez, M. Boguñá, A. Allard, and M. Á. Serrano, The hidden hyperbolic geometry of international trade: World trade atlas 1870–2013, *Sci. Rep.* **6**, 33441 (2016).
- [36] A. Allard and M. Á. Serrano, Navigable maps of structural brain networks across species, *PLoS Comput. Biol.* **16**, e1007584 (2020).
- [37] G. García-Pérez, M. Á. Serrano, and M. Boguñá, Soft communities in similarity space, *J. Stat. Phys.* **173**, 775 (2018).
- [38] K. Zuev, M. Boguñá, G. Bianconi, and D. Krioukov, Emergence of soft communities from geometric preferential attachment, *Sci. Rep.* **5**, 9421 (2015).
- [39] B. Klimt and Y. Yang, The enron corpus: A new dataset for email classification research, in *Proceedings of 15th European Conference on Machine Learning: ECML'04, Pisa, Italy* (Springer-Verlag, Berlin, Heidelberg, 2004), pp. 217–226.
- [40] K. Claffy, Y. Hyun, K. Keys, M. Fomenkov, and D. Krioukov, Internet mapping: From art to science, in *Proceedings of the 2009 Cybersecurity Applications & Technology Conference for Homeland Security, Washington, DC* (IEEE, Piscataway, NJ, 2009), pp. 205–211.
- [41] J. Serrà, A. Corral, M. Boguñá, M. Haro, and J. L. Arcos, Measuring the evolution of contemporary western popular music, *Sci. Rep.* **2**, 521 (2012).
- [42] R. Axelrod, *The Evolution of Cooperation* (Basic Books, New York, 1984).
- [43] M. Nowak and R. Highfield, *SuperCooperators: Altruism, Evolution, and Why We Need Each Other to Succeed* (Free Press, New York, 2011).
- [44] P. Higgs and N. Lehman, The RNA world: Molecular cooperation at the origins of life, *Nat. Rev. Genet.* **16**, 7 (2015).
- [45] M. A. Cant, The role of threats in animal cooperation, *Proc. Roy. Soc. B* **278**, 170 (2010).
- [46] S. Assenza, J. Gómez-Gardeñes, and V. Latora, Enhancement of cooperation in highly clustered scale-free networks, *Phys. Rev. E* **78**, 017101 (2008).
- [47] F. C. Santos, J. F. Rodrigues, and J. M. Pacheco, Graph topology plays a determinant role in the evolution of cooperation, *Proc. Roy. Soc. B* **273**, 51 (2006).
- [48] A. Szolnoki, M. Perc, and Z. Danku, Towards effective payoffs in the prisoner's dilemma game on scale-free networks, *Physica A* **387**, 2075 (2008).
- [49] F. C. Santos, J. M. Pacheco, and T. Lenaerts, Evolutionary dynamics of social dilemmas in structured heterogeneous populations, *Proc. Natl. Acad. Sci. USA* **103**, 3490 (2006).
- [50] J. Gómez-Gardeñes, M. Campillo, L. Floría, and Y. Moreno, Dynamical Organization of Cooperation in Complex Topologies, *Phys. Rev. Lett.* **98**, 108103 (2007).
- [51] C. Gracia-Lázaro, A. Ferrer, G. Ruiz, A. Tarancón, J. A. Cuesta, A. Sánchez, and Y. Moreno, Heterogeneous networks do not promote cooperation when humans play a prisoner's dilemma, *Proc. Natl. Acad. Sci. USA* **109**, 12922 (2012).
- [52] K.-K. Kleineberg, Metric clusters in evolutionary games on scale-free networks, *Nat. Commun.* **8**, 1888 (2017).
- [53] R. Cohen, K. Erez, D. ben Avraham, and S. Havlin, Resilience of the Internet to Random Breakdowns, *Phys. Rev. Lett.* **85**, 4626 (2000).
- [54] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Network Robustness and Fragility: Percolation on Random Graphs, *Phys. Rev. Lett.* **85**, 5468 (2000).
- [55] F. Nielsen, *Introduction to HPC with MPI for Data Science. (Chapter 8: Hierarchical Clustering)* (Springer, Switzerland, 2016).
- [56] C. Priebe, J. Conroy, D. Marchette, and Y. Park, Enron data set (2005).
- [57] F. Papadopoulos, M. Kitsak, M. A. Serrano, M. Boguna, and D. Krioukov, Popularity versus similarity in growing networks, *Nature (London)* **489**, 537 (2012).
- [58] M. Starnini, E. Ortiz, and M. Á. Serrano, Geometric randomization of real networks with prescribed degree sequence, *New J. Phys.* **21**, 053039 (2019).
- [59] R. Cressman and Y. Tao, The replicator equation and other game dynamics, *Proc. Natl. Acad. Sci. USA* **111**, 10810 (2014).
- [60] D. Helbing, A stochastic behavioral model and a microscopic foundation of evolutionary game theory, *Theor. Decis.* **40**, 149 (1996).
- [61] J. Poncela, *Evolutionary Games in Complex Topologies: Interplay Between Structure and Dynamics* (Springer Science and Business Media, Heidelberg, Germany, 2012).