Random walks on hypergraphs

Timoteo Carletti,[1] Federico Battiston,[2] Giulia Cencetti,[3] and Duccio Fanelli[4]

[1]*Namur Institute for Complex Systems, University of Namur, 5000 Namur, Belgium*
[2]*Department of Network and Data Science, Central European University, Budapest 1051, Hungary*
[3]*Mobile and Social Computing Lab, Fondazione Bruno Kessler, Via Sommarive 18, 38123 Povo, Trento, Italy*
[4]*Dipartimento di Fisica e Astronomia, Università di Firenze, INFN, and CSDC, Via Sansone 1, 50019 Sesto Fiorentino, Firenze, Italy*

(Received 14 November 2019; accepted 20 January 2020; published 18 February 2020)

In the past 20 years network science has proven its strength in modeling many real-world interacting systems as generic agents, the nodes, connected by pairwise edges. Nevertheless, in many relevant cases, interactions are not pairwise but involve larger sets of nodes at a time. These systems are thus better described in the framework of hypergraphs, whose hyperedges effectively account for multibody interactions. Here we propose and study a class of random walks defined on such higher-order structures and grounded on a microscopic physical model where multibody proximity is associated with highly probable exchanges among agents belonging to the same hyperedge. We provide an analytical characterization of the process, deriving a general solution for the stationary distribution of the walkers. The dynamics is ultimately driven by a generalized random-walk Laplace operator that reduces to the standard random-walk Laplacian when all the hyperedges have size 2 and are thus meant to describe pairwise couplings. We illustrate our results on synthetic models for which we have full control of the high-order structures and on real-world networks where higher-order interactions are at play. As the first application of the method, we compare the behavior of random walkers on hypergraphs to that of traditional random walkers on the corresponding projected networks, drawing interesting conclusions on node rankings in collaboration networks. As the second application, we show how information derived from the random walk on hypergraphs can be successfully used for classification tasks involving objects with several features, each one represented by a hyperedge. Taken together, our work contributes to unraveling the effect of higher-order interactions on diffusive processes in higher-order networks, shedding light on mechanisms at the heart of biased information spreading in complex networked systems.

DOI: 10.1103/PhysRevE.101.022308

I. INTRODUCTION

From social systems and the World Wide Web to economics and biology, networks define a powerful tool to describe many real-world systems [1–3]. Over the past 20 years of network science [4,5], many interacting systems with different functions were shown to exhibit surprisingly similar structural properties, at different scales. Interestingly, the complex architecture of real-world networks was found to significantly interfere with the dynamical processes hosted on them, from social dynamics [6] to synchronization [7]. As a consequence, properly tailored dynamical processes are now routinely employed to extract information on the *a priori* unknown structure of the underlying graphs architectures.

Networks materialize as pairwise interactions, represented by edges, among generic agents, the nodes: By their very first definition they are thus bound to encode binary relationships among units. However, an increasing amount of data indicates that, from biological to social systems, real-world interactions often occur among more than two nodes at a time. This phenomenon is not properly described by the traditional paradigm constrained on pairwise interactions and highlights the need for extended notions in the realm of network theory. In recent years, an emerging stream of research has been focusing on developing higher-order network models that account for diverse kinds of higher-order dependences, as found in complex systems.

Let us here observe that the current high-order framework bears some ambiguity, as it has been occasionally assumed to embrace features which are more specifically stemming from interactions [8], e.g., temporal and/or memory effects [9,10], or reflect the multiplex nature of the examined system [11–13]. Here the term "higher order" is exclusively meant to refer to agents interacting in groups of arbitrary numerosity [14–17], a process often modeled via simplicial complexes [18–20] or hypergraphs [21–23], and nontrivial mathematical generalization of the ordinary networks.

Our focus is on hypergraphs, where relationships among agents are described as collections of nodes assembled in sets, called hyperedges, made by any number of nodes. Hypergraphs provide a natural representation for many higher-order real-world networks [20,24]. In social systems they can, for instance, be suited to describe collaboration networks, where nodes denote authors and hyperedges stand for groups of authors, who have written papers together. Alternatively, hypergraphs can be invoked to describe face-to-face social networks where individuals can interact in groups of arbitrary sizes [25]. In biology, hypergraphs allow one to properly model biochemical reactions simultaneously involving more than two species or conveniently describe higher-order interactions among different families of proteins [15]. Crucially, in all these examples, interactions among agents occur in groups of arbitrary size and cannot be split into disjoint

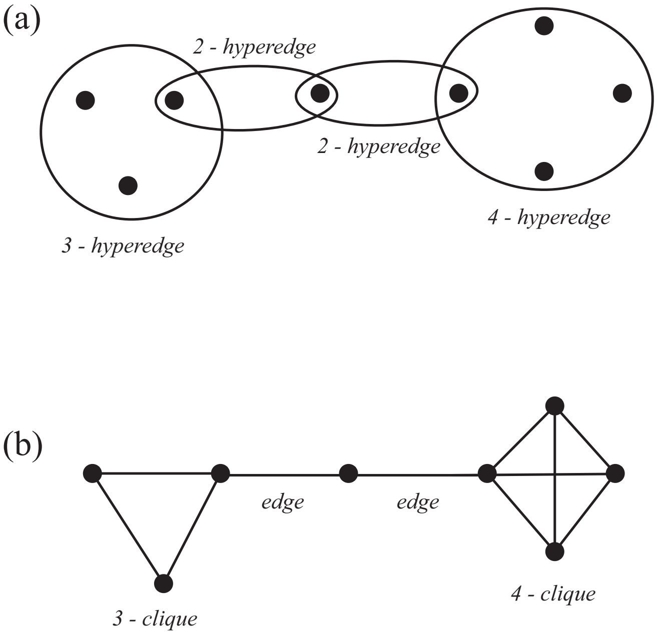

FIG. 1. (a) Hypergraph and (b)corresponding projected network. In the projected network each hyperedge E_α becomes a complete clique of size $|E_\alpha|$, thus with $|E_\alpha|(|E_\alpha|-1)/2$ pairwise interactions.

pairwise interactions. Differently from simplicial complexes, a higher-order interaction described by a hypergraph, e.g., a single three-body interaction, does not require the existence of all lower-order interactions, e.g., the three pairwise interactions associated with the same triangle [19]. Heterogeneous hypergraphs have been sometimes studied by mapping the nodes belonging to a hyperedge into a clique of suitable size. However, the drawback of this procedure is that it eventually yields a *projected* network, e.g., shown in Fig. 1, where only pairwise interactions are ultimately accounted for (see Appendix A).

Linear dynamics [26–28] and specifically random walks [29] constitute a simple although powerful tool to extract information on the relational structure of interacting systems. In particular, random walks on complex networks [30] have been proven useful to compute centrality scores [31], finding communities [32] and providing a taxonomy of real-world networks [33]. In the simplest case, at each time step, a walker jumps from the node where it belongs to one of its adjacent neighbors, traveling across one of the available edges, chosen at random with uniform probability. Many variations of this fundamental process have been considered. These include more sophisticated dynamical implementations, which allow one to target the walks towards nodes with given structural features [34], let them interact at the nodes of the network [35], investigate nonlinear transition probabilities [36] and crowded conditions [37], and consider the temporal [38–40] or multilayer [41,42] dimensions of the edges under different network topologies.

Random walks have been defined on simplicial complexes [43], but because of the cumbersome combinatorics involved, applications have been limited to higher-order interactions of the lowest dimensions, i.e., triangles. Moreover, walkers are in general allowed to hop between edges or even high-order structures. This is at variance with the setting that we here aim at exploring, where hops can solely occur among nodes which join in a given high-order structure. In parallel, also random walks on a hypergraph have been considered by assuming that all the hyperedges are made by an identical (and constant) number of nodes [44,45]. The first random-walk Laplacian defined on hypergraphs can be probably traced back to the seminal paper by Zhou *et al.* [46]. Each hyperedge is endowed with an arbitrary weight, acting as a veritable bias to the walker dynamics. As observed by the authors of [46], assigning the weights is an outstanding open problem, which deserves to be properly addressed. In this work, we will prove that a physically motivated choice for the aforementioned weights naturally emerges when framing the problem on solid microscopic grounds.

Interestingly, more complicated nonlinear dynamics have also been recently studied on simplicial complexes [16,20,47,48] or in a pure multibody frame [49]. Once again, however, the focus is placed on low-dimensional simplicial complexes (triangles). Recently, several dynamics, including epidemic spreading [16,47,50] and synchronization [36], have been shown to produce new collective behaviors when higher-order interactions are assumed to shape the networked arrangement.

Starting from this setting and elaborating on the above, we propose in this work a class of random walks evolving on generic heterogeneous hypergraphs as dictated by a plausible physical model and without any limitation on the sizes of the hyperedges. In this framework, multibody proximity is associated with highly probable exchanges among agents belonging to the same hyperedge, and walkers mitigate their inclination to explore the system with a tendency to naturally spend more time in highly clustered cliques and communities. This feature is reminiscent of bias in information spreading, which is known to be affected by the phenomenon of echo chambers [51]. Similarly to the standard random walk, at each time step a walker sitting on a node selects a node from its neighborhood, i.e., a node belonging to the one of the hyperedges where it happens to be, with a probability of jump weighting the size of the hyperedges and taking into account the number of hyperedges to which the selected node belongs. In this way, higher-order interactions between a group of nodes drive the process and the weights postulated in [46] take nontrivial values, as stemming from the microscopic dynamics.

We will in particular provide an analytical description of the process, by deriving a general formula for the stationary distribution of the walk, and show that the dynamics is driven by a generalized Laplace operator that reduces to the standard random-walk Laplacian when all hyperedges have size 2 and the hypergraph results in a traditional network.

As already stated, random walks can be used to rank nodes, based on the stationary occupancy probability of walkers across the network. Because of these implications, it is therefore interesting to compare the stationary distribution, as obtained within the framework introduced herein, with that displayed by standard random walkers on the corresponding projected network. Because of the tight interactions among agents belonging to the same hyperedge, the probability to find a walker on a given node is in principle different when confronting the outcome of the two aforementioned processes. As a consequence, we expect a different order in the ranking to be obtained for the same node, depending on the dynamical process employed in the analysis. This observation opens up the possibility for an alternative definition of centrality for systems where the high-order structure is known to be relevant. In particular, we will provide direct evidence for our claims working with coauthorship networks, as extracted from the arXiv online preprint server. Our second application, to which we alluded above, concerns a classification task which is borrowed from [46]. Indeed, it is well known that one can model a data set by resorting to networks and then make use of the associated Laplacian eigenmaps [52] to embed the data on a lower-dimensional space, while hopefully preserving relevant information, in the spirit of a generalized principal component analysis. Working in the lower-dimensional space allows one to cluster together objects. However, when objects to be classified share annotated features, the use of binary relationships, i.e., the usual network, results in a dramatic loss of information. One can thus obtain a better embedding via hypergraphs and invoke the spectral characteristics of the associated Laplacian to achieve more effective clustering scores [46,52]. Inspired by the analysis in [46], we will here consider the problem of separating items taken from different databases in the UCI Machine Learning Depository in distinct classes, by using within the scope a set of annotated features. Hence here nodes are items and hyperedges features. We will show that the presence of high-order interactions among features as encoded via the proposed Laplacian operator yields a very effective embedding with just a few of the most significative directions. Our results are in line (and in some cases more accurate) than those reported in [46].

Summing up, we here introduce and discuss a generalization of the random-walk picture to higher-order networked systems, where hyperedge weights are naturally assigned, thus removing any ambiguity in their values. Finally, we hint at important exploitations of this dynamical framework working along two paradigmatic directions, ranking, and classification of data.

II. MODEL

A. Incidence and adjacency matrices of the hypergraph

Let us consider a hypergraph $\mathcal{H}(V, E)$, where $V = \{1, \dots, n\}$ is the set of n nodes and $E = \{E_1, \dots, E_m\}$ the set of m hyperedges, with E_α an unordered collection of nodes, i.e., $E_\alpha \subset V \,\forall\, \alpha = 1, \dots, m$. We observe that whenever $E_\alpha = \{i, j\}$, i.e., $|E_\alpha| = 2$, the hyperedge is actually a "standard" edge, denoting a binary interaction among nodes i and j. A hypergraph where $|E_\alpha| = 2 \,\forall\, \alpha$ reduces to a network.

We can define the associated incidence matrix of the hypergraph $e_{i\alpha}$, carrying the information about how nodes are shared among hyperedges, as

$$e_{i\alpha} = \begin{cases} 1 & \text{for } i \in E_\alpha \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

We note that the same matrix exists for networks. However, while in regular networks each column can have only two nonzero entries, as each edge can contain two nodes only,[1] in hypergraphs each column can display several nonzero entries, i.e., a hyperedge can contain several nodes.

Starting from the above matrix, one can construct the $n \times n$ adjacency matrix of the hypergraph $A = ee^T$, whose entry A_{ij} represents the number of hyperedges containing both nodes i and j. We note that often the adjacency matrix is defined by setting to 0 the main diagonal. Let us also define the $m \times m$ hyperedges matrix $C = e^T e$, whose entry $C_{\alpha\beta}$ counts the number of nodes in $E_\alpha \cap E_\beta$. Observe that in the literature the number of nodes in a given hyperedge $C_{\alpha\alpha}$ is often called the degree of the hyperedge, while the node degree stands for the number of hyperedges containing the node $\sum_\alpha e_{i\alpha} e_{i\alpha}$.

B. Transition probability

To describe a random-walk process, we need to define the transition probability to pass from a state, here represented by the node on which the walker belongs, to any other state, compatible with the former, in one time step. In the case of simple unbiased random walks on networks, one assumes the walker to take with equal probability any link emerging from the node that is initially occupied. Hence, the transition probability can be readily computed as A_{ij}/k_i, where $k_i = \sum_j A_{ij}$ is the degree of the origin node. When dealing with hypergraphs, choosing with uniform probability any of the neighboring nodes, namely, all the nodes belonging to hyperedges connected with the origin node, is not a sensible choice. In this way, in fact, the real structure of the systems is not incorporated into the dynamical picture. On the contrary, nodes belonging to the same hyperedge exhibit a higher-order interaction and we consequently assume that spreading among them is more probable than with nodes associated with other hyperedges; because of this, the information can thus spend long periods inside the same hyperedge. For instance, gossip can spread faster, because of group interaction among individuals, than as follows successive binary encounters; similarly, ideas can circulate more effectively among collaborators, the coauthors of a joined publication, as compared to the setting where exchanges in pairs are solely allowed for. Roughly speaking a walker sitting on a node "assigns" to all its neighbors a weight that senses the size of the hyperedges and corrects for the number of shared hyperedges. Thus, to compute the transition probability to jump from i to j, we count the number of nodes, excluding i itself, belonging to the same hyperedge of i and j. Recalling the definition of the matrix C, this can be written as

$$k_{ij}^H = \sum_\alpha (C_{\alpha\alpha} - 1) e_{i\alpha} e_{j\alpha} = (e\hat{C}e^T)_{ij} - A_{ij} \,\forall\, i \neq j \tag{2}$$

and $k_{ii}^H = 0$, where $\hat{C}$ is a matrix whose diagonal coincides with that of C and it is zero otherwise (see Appendix B). By normalizing so as to impose a uniform choice among the connected hyperedges, we get the expression for the transition

[1] We do not consider here hyperedges with size 1 because they correspond to isolated nodes, i.e., nodes that cannot take part in the examined process.

probabilities

$$T_{ij} = \frac{(e\hat{C}e^T)_{ij} - A_{ij}}{\sum_{\ell\neq i} k^H_{i\ell}} = \frac{(e\hat{C}e^T)_{ij} - A_{ij}}{\sum_{\ell\neq i}(e\hat{C}e^T)_{i\ell} - k^H_i}, \tag{3}$$

where $k_i^H = \sum_{\ell\neq i} A_{i\ell}$ is the hyperdegree of the node i, a measure reminiscent of the node degree, which takes into account both the number and the size of hyperedges in which i is involved.

When the hypergraph is a network, all hyperedges have two nodes. Hence

$$(e\hat{C}e^T)_{ij} = \sum_\alpha C_{\alpha\alpha} e_{i\alpha} e_{j\alpha} = 2\sum_\alpha e_{i\alpha} e_{j\alpha} = 2A_{ij} \tag{4}$$

and Eq. (3) reduces to the standard transition probability for a random walk on networks

$$T_{ij} = \frac{2A_{ij} - A_{ij}}{2k_i^H - k_i^H} = \frac{A_{ij}}{k_i}, \tag{5}$$

where we used the fact that, under this assumption, $k_i^H = k_i$.

Remark. For a lazy random walk we observe that one can straightforwardly generalize the above mechanism as to include the possibility for the walker to remain on the same node

$$T_{ij}^{(\text{lazy})} = \frac{k_{ij}^{(\text{lazy}),\text{H}}}{\sum_\ell k_{i\ell}^{(\text{lazy}),\text{H}}}, \tag{6}$$

where now

$$k_{ij}^{(\text{lazy}),\text{H}} = \sum_\alpha C_{\alpha\alpha} e_{i\alpha} e_{j\alpha}\, \forall\, i,\, j. \tag{7}$$

C. Stationary solution

Having computed the transition probabilities, we can proceed further by formulating the dynamical equation which rules the temporal evolution of the probability $\mathbf{p}(t) = (p_1(t), \dots, p_n(t))$ of finding the walker on a given node after $t > 0$ steps. The process is governed by the equation

$$p_i(t+1) = \sum_j p_j(t) T_{ji}, \tag{8}$$

where the term on the right-hand side combines the probability to be in any node j at time t and the probability to perform a jump towards the target node i, during the next time of iteration. As $\sum_j T_{ij} = 1$ for all i, the stationary probability distribution $\mathbf{p}^{(\infty)}$ is thus the left eigenvector associated with the eigenvalue $\lambda_1 = 1$ of $\mathbf{T}$.

Given $\mathbf{T}$, it is possible to obtain an exact analytical solution for the stationary state $\mathbf{p}^{(\infty)}$ which encapsulates the higher-order structure of the system. Indeed, a straightforward computation (see Appendix C) yields

$$p_j^{(\infty)} = \frac{\sum_{\ell\neq j}(e\hat{C}e^T)_{j\ell} - k_j^H}{\sum_{m\neq\ell}\left[(e\hat{C}e^T)_{m\ell} - k_m^H\right]} \tag{9}$$

for all $j = 1, \dots, n$. By defining $d_j^H = \sum_{\ell\neq j} k_{j\ell}^H$, one can rewrite the previous equation as $p_j^{(\infty)} = d_j^H / \sum_j d_j^H$, which is reminiscent of the typical expression for the stationary distribution for the random walk on networks. Indeed, when the hypergraph is a standard binary network $d_j^H = k_j$, i.e., the node degree, and we recover exactly the well-known expression $q_j^{(\infty)} = k_j / \sum_l k_l$ for the stationary solution of the walk.

We observe that

$$L_{ij} = \delta_{ij} - T_{ij} = \delta_{ij} - \frac{k_{ij}^H}{\sum_{\ell\neq i} k_{i\ell}^H} \tag{10}$$

is a different random-walk Laplacian that generalizes the random-walk Laplacian for networks. Moreover, the former reduces to the latter in the case $|E_\alpha| = 2$ for all α.

We observe that the formalism readily extends to the case of continuous-time random walks, where the evolution of the probability is given by

$$\dot{p}_i(t) = \sum_j p_j(t) T_{ji} - \sum_j p_i T_{ij}.$$

Similarly to the case of networks, such as $\sum_j T_{ij} = 1$, it is possible to rewrite the preceding equation as

$$\dot{p}_i = \sum_j p_j (T_{ji} - \delta_{ij}) = -\sum_j p_j L_{ji}\,,$$

where $\mathbf{L}$ is the above-defined Laplace matrix. In the following, for the sake of definiteness, we limit our analysis to exploring the properties of discrete-time random walks on synthetic and real-world hypergraphs, leaving the continuous-time case to future work.

We denote by $\mathbf{D}$ the diagonal matrix with entries $d_i^H = \sum_{j\neq i} k_{ij}^H$ and by $\mathbf{K}^H$ the matrix characterized by elements k_{ij}^H. We can introduce the symmetric Laplacian $\mathbf{L}^{\text{sym}}$ as

$$\mathbf{L}^{\text{sym}} = \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{K}^H\mathbf{D}^{-1/2}\,,$$

which is well defined since $k_{ij}^H \geqslant 0$. Here $\mathbf{L}^{\text{sym}}$ is similar to the operator introduced via the relation (10); indeed, $\mathbf{L} = \mathbf{D}^{-1/2}\mathbf{L}^{\text{sym}}\mathbf{D}^{1/2}$. The operator $\mathbf{L}$ introduced herein is hence a properly defined Laplacian: It is in fact non-negative definite; it displays real eigenvalues and the smallest eigenvalue is identically equal to zero, as it readily follows by virtue of the proven similarity to $\mathbf{L}^{\text{sym}}$.

Before turning to discussing the applications, we will briefly draw a comparison with the setting proposed by Zhou *et al.* [46] and show how this materializes in a natural solution for the problem of weight determination. The Laplacian operator $\mathbf{L}^z$ introduced in [46] removing the possibility for the walker to stay put on the node can be cast in the form

$$L_{ij}^z = \delta_{ij} - \sum_\alpha \frac{w_\alpha}{W_i(C_{\alpha\alpha} - 1)} e_{i\alpha} e_{j\alpha}, \tag{11}$$

where w_α identifies the arbitrary weight of the hyperedge E_α, $W_i = \sum_\alpha w_\alpha e_{i\alpha}$ is the total weight of the hyperedges containing the node i, i.e., weighted node degree, and $C_{\alpha\alpha}$ stands for the number of nodes in the hyperedge E_α. A simple calculation, as detailed in the following, shows that the operator $\mathbf{L}$ can be eventually recovered from $\mathbf{L}^z$ by imposing

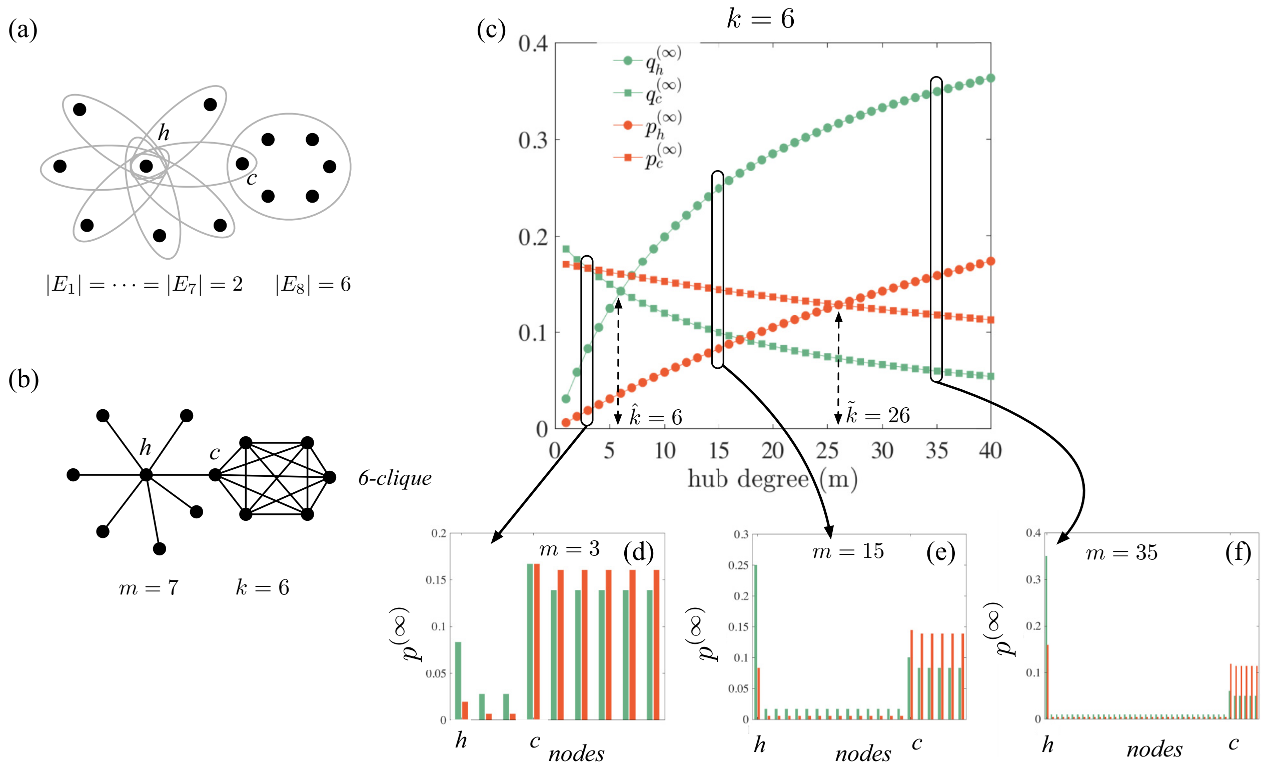

FIG. 2. The (m, k)-star-clique network. (a) Hypergraph made by $m + k = 13$ nodes, divided into $m = 7$ hyperedges of size 2 and one large hyperedge of size $k = 6$. The node h belongs to all the 2-hyperedges, while the node c belongs to one 2-hyperedge and to the 6-hyperedge. (b) Projected network where hyperedges are mapped into complete cliques, and the 6-hyperedge becomes thus a 6-clique. (c) Dependence on m of the asymptotic probability of finding the walker on the node h (circles) or on the node c (squares), in the projected network (green symbols) and in the hypergraph (orange symbols). The asymptotic probabilities $q_i^{(\infty)}$ and $p_i^{(\infty)}$ are reported for three values of m: (d) $m = 3 < \hat{k}$, (e) $\hat{k} < m = 15 < \tilde{k}$, and (f) $\tilde{k} < m = 35$, where $\hat{k} = 6$ and $\tilde{k} = 26$.

the nontrivial weights $w_\alpha = (C_{\alpha\alpha} - 1)^2$. In fact,

$$\begin{aligned}
L_{ij}^z &= \delta_{ij} - \sum_\alpha \frac{(C_{\alpha\alpha} - 1)^2}{(C_{\alpha\alpha} - 1)\sum_\beta (C_{\beta\beta} - 1)^2 e_{i\beta}} e_{i\alpha} e_{j\alpha} \\
&= \delta_{ij} - \sum_\alpha \frac{C_{\alpha\alpha} - 1}{\sum_\beta (C_{\beta\beta} - 1)^2 e_{i\beta}} e_{i\alpha} e_{j\alpha} \\
&= \delta_{ij} - \frac{k_{ij}^H}{\sum_\beta \sum_{\ell \neq i} e_{\ell\beta} (C_{\beta\beta} - 1) e_{i\beta}} \\
&= \delta_{ij} - \frac{k_{ij}^H}{\sum_{\ell \neq i} k_{i\ell}^H} = L_{ij},
\end{aligned}$$

where use has been made of the definition (2) for k_{ij}^H and the fact that $C_{\beta\beta} - 1 = \sum_{\ell \neq i} e_{\ell\beta}$. As anticipated, a natural choice for the weights as postulated in [46] can be envisaged, which follows a sensible microscopic modeling of the random-walk dynamics. Let us observe that in the case of a lazy random walk, a similar result can be obtained by setting $w_\alpha = C_{\alpha\alpha}(C_{\alpha\alpha} - 1)$.

By invoking Theorem 4 in [53], we can finally conclude that our process is equivalent to a random walk on a weighted projected network, where the weight of the link ij is given by k_{ij}^H, that is, the weights scale extensively with the region of influence of the nodes, namely, the size of the hyperedge they belong to. It is indeed quite remarkable that a properly weighted binary network encapsulates the higher-order information, as stemming from the corresponding hypergraph representation. Observe that authors in [53] also consider an extension of the Zhou *et al.* model, where nodes bear a given weight, tuned so as to reflect the hyperedge characteristics. Again, the introduced weights are abstract quantities and do not reflect a physically motivated choice.

III. RESULTS

Since the creation of the PageRank algorithm [54,55], random walks on networks have been routinely applied to compute centrality scores [31]. Indeed, they can be used to rank nodes according to the probability to be visited by the walker; the larger the walker, the more important or central the node. In this section we show that high-order interactions can strongly modify the ranking, as resulting from a random-walk process on hypergraphs, with respect to the homologous estimate as computed for the corresponding projected network. This fact can thus bear relevant implications for ranking real data, stemming from a dynamical process which is better explained in terms of hypergraphs. In this case, in fact, the applications of ranking tools tailored to pairwise interactions might produce misleading results (see Appendixes C and E).

To illustrate the effect of a nontrivial higher-order structure, we consider a simple hypergraph made by m hyperedges of size 2 all intersecting in a common node h; a different node,

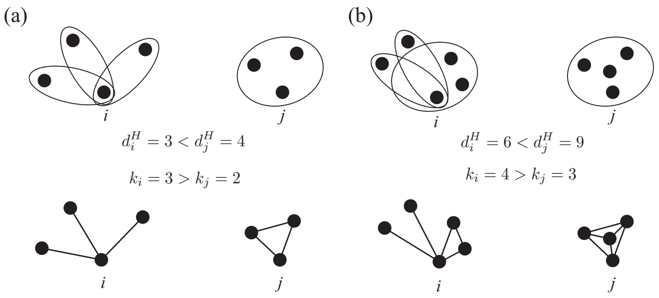

FIG. 3. Examples with ranking inversion. We propose two typical examples of high-order structures that locally produce two different rankings: (a) an example involving three 2-hyperedges and one 3-hyperedge and (b) the case with one 3-hyperedge and two 2-hyperedges compared with a 4-hyperedge. In both cases the first configuration will be ranked above the second one, when using the random walks on hypergraphs, while the opposite holds when the random walkers run on the projected networks.

say, c, belongs to one such 2-hyperedges and to a hyperedge of size k [see Fig. 2(a) for the case $m = 7$ and $k = 6$].

The random walk on the projected network will rank nodes according to their degree, i.e., $q_i^{(\infty)} \sim k_i$. Hence, for $m > k$, the node h, with $k_h = m$, is ranked first, followed by the c node, $k_c = k$, and all others [see the green curves in Fig. 2(c)]. In contrast, the random walk on the hypergraph ranks nodes taking into account higher-order relations. Since from Eq. (9) we get $p_h^{(\infty)} \sim m$ and $p_c^{(\infty)} \sim 1 + (k-1)^2$, h is the top node as long as $m > 1 + (k-1)^2$ [see the orange curves in Fig. 2(c)]. In conclusion, for a fixed size of the hyperedge k, if the "hub" node is too small [see Fig. 2(d)] $m < \hat{k} = k + 1$ or the hub is very large [see Fig. 2(f)] $m > \tilde{k} = 1 + (k-1)^2$, then both processes will rank nodes in the same way. However, there exists a range of intermediate values $\hat{k} < m < \tilde{k}$ for which the top ranked node on the hypergraph is the c node while the random walk on the projected network returns the h node as the top rank [see Fig. 2(e)].

This phenomenon of ranking inversion can be roughly stated as follows: With the aim of maximizing the probability of occupancy of a given node, it is preferable for this latter to be connected to nodes organized into a few large hyperedges than many parceled units. More precisely, the analytical expression for $\mathbf{p}^{(\infty)}$ indicates that the ranking provided by the random walkers on the hypergraph is proportional to $d_i^H = \sum_j k_{ij}^H$, while it is well known that the ranking that follows the usual random walks on the projected network scales proportionally to the node degree k_i. Two nodes, say, i and j, are thus ranked differently by the two processes, if $k_i > k_j$ but $d_i^H < d_j^H$. As we will now show, the presence of high-order structures can induce a ranking inversion.

A simple example where this occurs is shown in Fig. 3(a). The node i belongs to the intersection of three 2-hyperedges. Thus its degree (in the projected network) is given by $k_i = 3$. Moreover, $d_i^H = 3$, because, locally, the hypergraph reduces to a standard network; on the other hand, the node j belongs to a 3-hyperedge, and hence $k_j = 2$, because it is part of a 3-clique, but $d_j^H = 4$. Hence, $k_i > k_j$ but $d_i^H < d_j^H$. Nodes j will be consequently ranked above i using the generalized random walks on the hypergraph, while the opposite happens if one relies on random walks on the projected network.

The above construction can be readily generalized, as shown by the example presented in Fig. 3(b). Here i belongs to a 3-hyperedge and to two 2-hyperedges, and hence $k_i = 4$ and $d_i^H = 6$; node j instead belongs to a 4-hyperedge, and thus $k_j = 3$ and $d_j^H = 9$. So again $k_i > k_j$ while $d_i^H < d_j^H$.

To further characterize the impact of the high-order interactions on diffusion on larger systems, we consider a second synthetic model where all nodes have the same number of neighbors, which are arranged in a tunable number of triangles, i.e., hyperedges of size $|E_\alpha| = 3$. The model interpolates between the case where the number of triangles is zero $f = 0$, meaning that all interactions involve simple pairs, and the case where there are no pairwise interactions but only three-body ones, $f = 1$. More precisely, we start with a one-dimensional (1D) regular lattice where nodes are connected to four neighbors (two on the left and two on the right). Each nodes hence has degree 4 and takes part in two distinct triangles, i.e., hyperedges with size 3, and $f = 1$. Then, with probability p, we iteratively swap the ending points of the links with a crisscross rewiring, i.e., preserving the node degree, progressively eliminating 3-hyperedges, hence triangles. In the limit of high rewiring triangles have a negligible probability to be formed and one eventually obtains a regular random graph with degree $k = 4$. In the process, we control that no hyperedge of size greater than 3 is created, so the competition is only between two-body and three-body interactions.

As the degree sequence is unchanged throughout this process and every node shares the same number of links, the asymptotic distribution of walkers on the projected network is uniform and given by $q_i = 1/N$ for all i, where N is the number of nodes, set to 500 in the example below, no matter the value of f. This is also the case for the random walk on the hypergraph, in the two limiting cases $f = 0$ and $f = 1$; indeed, in the former case the hypergraph and the projected network coincide because all the hyperedges have size 2. In the latter setting, all nodes are involved in the same number of higher-order interactions and thus they are all equivalent. However, for the walk on hypergraphs the stationary state changes at the intermediate stages of f. In order to quantify the heterogeneity of the stationary state we rely on the Gini coefficient, which is defined as the average absolute difference between all pairs of elements in the vector p, divided by the average

$$G(p) = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} |p_i - p_j|}{2N\sum_{i=1}^{N} p_i}. \tag{12}$$

The Gini coefficient for the stationary state of a random walk on the above-described hypergraph is reported in Fig. 4(a). For the limiting values $f = 0$ and $f = 1$ the stationary state on the hypergraph coincides with the one on the projected network and the Gini index is 0 since the asymptotic solution is homogeneous. However, high-order structures arising for intermediate values of the fraction of triangles induce a heterogeneity in the occupation of the different nodes at equilibrium, which is thus different from the one obtained for the associated projected network.

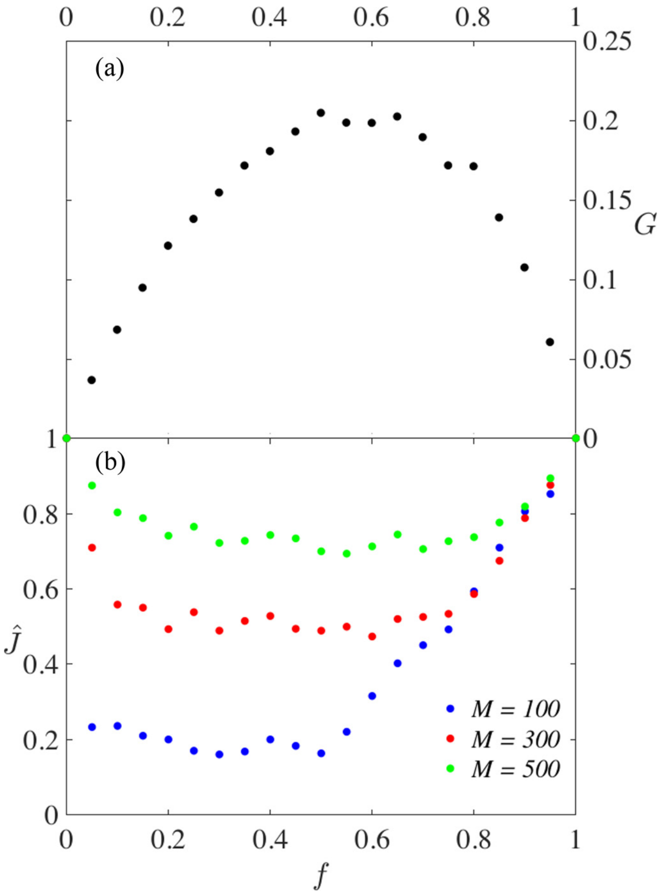

FIG. 4. Impact of the three-body interaction on the asymptotic solution of the random walk on the hypergraph. (a) Gini coefficient for the stationary state of the random walk on hypergraphs as a function of the fraction of hyperedges of size three, f. Recall that the model does not allow for hyperedges of size larger than 3. (b) For the same networks as in (a), the modified Jaccard index to compare the rankings of nodes for the hypergraph and the projected network. Different colors (blue, red, and green), correspond to different numbers of nodes chosen for the comparison, i.e., the top 100, the top 300, and all 500 nodes, respectively.

A standard metric to compare lists is the Jaccard index, a measure of the fraction of elements that are common between two lists with respect to the total number of involved elements $J(A, B) = |A \cap B|/|A \cup B|$. As the Jaccard index does not take into account the order of the elements as appearing in the two confronted lists, we compare the rankings of the two stationary distributions by means of a modified Jaccard index $\hat{J}$, recently introduced in [56]. Here differences at the top of the ranking induce a stronger change than differences associated with the lower-ranked elements. Let us observe also that the Jaccard index is unable to detect a permutation in the order of the elements on a list, while the modified one does. In Fig. 4(b) we show the average modified Jaccard index $\hat{J}$ for the M-top ranking, $M = 100, 300, 500$, as a function of the fraction f of 3-hyperedges existing in the system. The results are in agreement with the ones obtained via the Gini coefficient; for $f = 0$ and $f = 1$ the rankings coincide and thus $\hat{J}$ achieves its maximum value, i.e., 1, while for intermediate values of f the index $\hat{J}$ drops down, reflecting differences among the rankings. Moreover, we can appreciate the presence of a large turnover in the top lists: Indeed, the $\hat{J}$ associated with small M, i.e., comparing relatively few nodes in the top list, is much smaller than that for large M, i.e., longer lists.

To take one step forward, we consider a synthetic model where high-order structures are not limited to three-body interactions, but larger hyperedges are allowed for. We thus build a third model which interpolates from a 1D ring to a fully connected network. More precisely, we start from a 1D ring where all the nodes have degree 2 and then progressively increase its density as measured by the total number of links ℓ until the process terminates with a complete network, corresponding to a hypergraph with a single hyperedge containing all the nodes. Links are added at random avoiding self-loops and multiple links. We note that, differently from the previous case, at intermediate values of ℓ this model presents a much wider variety in the size of the hyperedges (or cliques in the projected network), which are no longer limited to two-body and three-body interactions. For this reason, the structure of the ranking difference is definitely more complex and rich than what one could eventually guess by just looking at the number of 3-hyperedges, 4-hyperedges, or 5-hyperedges (see Fig. 5).

In the initial configuration of a 1D ring, the stationary solutions of the hypergraph and the projected network coincide, because of the absence of higher-order interactions. Similarly, they are also equivalent in the opposite limit, i.e., when the fully connected network is generated. For an intermediate number of added links, the two processes result instead in different rankings. In Fig. 5 we report $\hat{J}$ as a function of the total number of links ℓ, to compare the M-top rankings, as obtained by using the random walk on the hypergraph and on the projected network, respectively. We report in particular results for three values $M = 5, 10, 20$. The behavior of the threes curves is qualitatively similar. Indeed, they all reach the value 1, i.e., perfect matching of the respective rankings for $\ell = 20$ (initial 1D ring). Then, even the addition of just a few links makes the rankings change abruptly and $\hat{J}$ consequently drops to low values. This is associated with the creation of small hyperedges with size equal to 3 [see Fig. 5(b)]. Adding more links reduces the differences, namely, $\hat{J}$ increases, up to $\ell = 190$ (complete network), where again the rankings coincide and the index equals 1. This is associated with the birth of larger hyperedges. We note that $\hat{J}$ for $M = 5$ is much smaller than the same quantity computed with $M = 10$ (rank half of the nodes) and $M = 20$ (rank all the nodes), meaning that there is a strong turnover in the top positions. The heterogeneity in the stationary solutions of this model, as well as the star-clique example, is further investigated in Appendix D, where the corresponding Gini coefficients are shown.

IV. APPLICATIONS

A. Node ranking

In the preceding section we showed that the hypergraph and the projected network can exhibit different stationary solutions because of ranking inversion (see previous discussion). We thus decided to analyze the impact of this observation in real networks of scientific collaborations, in our opinion one

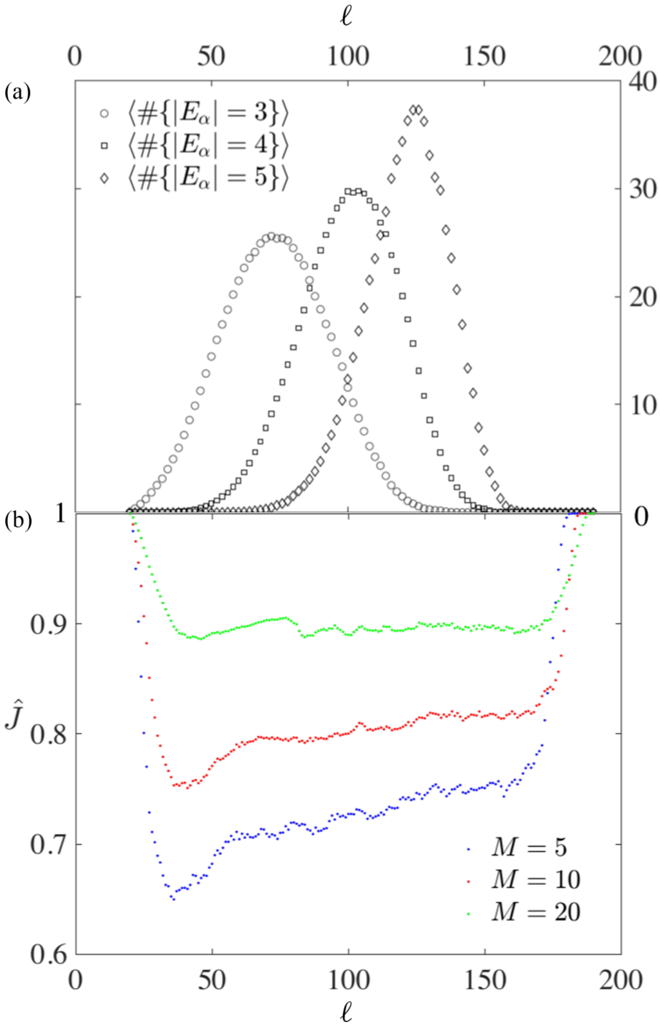

FIG. 5. Impact of high-order structures on the asymptotic distribution of walkers for the random walk on (a) the hypergraph and (b) the projected network. Using the algorithm presented in the text, by iteratively adding links we create hypergraphs that interpolate from a regular 1D ring (where each of the $N = 20$ nodes is connected with its two neighbors) to a complete graph. We then perform the random-walk process on the hypergraph and on the associated projected network and compare the resulting ranking (the top 5 blue, the top 10 red, and the top 20 green, i.e., the whole set of nodes) using $\hat{J}$ [in (b)]. For a small number of available links ℓ, the hypergraph does not present many hyperedges and thus the ranking is very close, $\hat{J} \sim 1$. As ℓ starts to increase, a few hyperedges of size 3 are created [see the circles in (a)] and the rankings estimated with the two alternative methods deviate, the values of $\hat{J}$ dropping in turn. However, as ℓ increases even more, larger high-order structures, e.g., 4- and 5-hyperedges, emerge [see squares and diamonds in (a)] and $\hat{J}$ steadily increases. For a large ensemble of added links $\ell \gtrsim 170$, the rankings become similar and $\hat{J} \sim 1$.

of the most representative examples of high-order structures in human interactions. The analyzed data have been gathered from the arXiv database (see Appendix E for more details). Human collaborations are often schematized as resorting to pairwise interaction, a working ansatz which amounts to ignoring the organization in teams. At variance, we have instead built a hypergraph where researchers (i.e., nodes) coauthoring an article are part of the same hyperedge.

We have determined the largest connected component of the hypergraph and that of the projected network, considered maximal and unique hyperedges (to have a fair comparison with the cliques), and computed (i) the stationary distribution $\mathbf{p}^{(\infty)}$ for the random walk on the associated hypergraph and (ii) the stationary distribution $\mathbf{q}^{(\infty)}$ for the random walk on the corresponding projected network. We then normalize the computed stationary probabilities by their relative maximum so as to favor a comparative visualization. In Fig. 6 we plot $p_i^{(\infty)}/\max_j p_j^{(\infty)}$ vs $q_i^{(\infty)}/\max_j q_j^{(\infty)}$ for the case of arXiv-astro and arXiv-physics. In Fig. 15 the same comparison is drawn for the complete arXiv data set.

Authors are ranked differently, according to the two criteria, the one based on hypergraphs being more sensitive to the organization in groups. If the computed rankings were (almost) the same, the data would (almost) lie on the main diagonal; deviation from this results in novel information conveyed by the random walk on the hypergraph. The unitary square in the plane $(q_i^{(\infty)}, p_i^{(\infty)})$ can be divided into four smaller squares (see Fig. 6). The majority of the authors lie in the bottom left square $[0, 1/2] \times [0, 1/2]$: These authors have therefore written a few papers with a small number of coauthors. Three other regions can however be identified which roughly correspond to the bounded squares: $[1/2, 1] \times [0, 1/2]$ (bottom right), $[0, 1/2] \times [1/2, 1]$ (top left), and $[1/2, 1] \times [1/2, 1]$ (top right). Authors in the top right square are top ranked in both processes; they have hence written a large number of papers with different collaborators (large degree), but they have also contributed to a relevant number of papers with many coauthors, i.e., large hyperedge size. Scholars in the bottom right square are better ranked by the random walk on the network; this means that they have written several papers but with a small number of coauthors [see, e.g., Fig. 6(b), corresponding to physics]. Finally, researchers in the top left square manifest a complementary attitude; they have participated in a small number of papers, but written by many authors [see e.g., Fig. 6(a), corresponding to astro].

As a further consideration, we can bring to the fore different "habits" of publication and writing papers that authors exhibit in each domain, despite the distribution of node degrees, i.e., number of different collaborators per author, and of hyperedges size, i.e., number of coauthors in papers, showing a quite similar shape across domains, e.g., broad tails (see Appendix E). This is particularly relevant for the high-energy particle (hep) archive, one among the oldest ones and divided into four subcategories: experimental (ex), lattice (lat), phenomenology (ph), and theory (th) (see Fig. 7). Indeed, hep-ex and hep-ph populate mainly the top right square, while hep-lat and hep-th are more present in the top right and bottom right squares. Researchers belonging to the former community therefore tend to write several papers with many coauthors, while those associated with the latter have papers with many different collaborators, each one coauthored by a small number of scholars. This is also confirmed by the largest degree found in the four subcategories (see Table I), which is as large as ∼1200 for hep-ex and hep-ph, while it is almost one-fourth the size for hep-lat and hep-th.

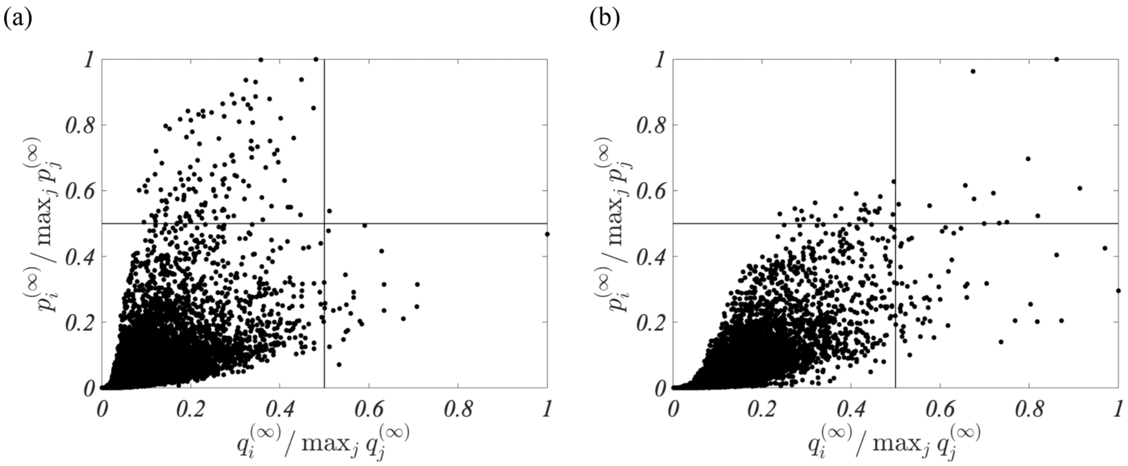

FIG. 6. Comparison of the rankings in the arXiv community for the case of (a) astro and (b) physics. We report the scatter plot of the normalized rankings obtained with the random walk on network $q_i^{(\infty)}$ and the one computed using the random walk on hypergraphs $p_i^{(\infty)}$ for (a) arXiv-astro and (b) arXiv-physics.

The comparison drawn may allow one to introduce apt corrections to usual bibliographic indicators, by properly weighting the participation to large collaborations, as opposed to research activities carried out in small groups. Recall that the hypergraph Laplacian is equivalent to the Laplacian obtained from a properly weighted projected network, which inherits the high-order structures of the hypergraph [53]. Assessing the higher-order ranking therefore amounts to applying the usual tools to this latter weighted binary graph, a conclusion which points to an immediate operative translation of the methods introduced herein.

B. Classification task

To further test the interest of a generalized random-walk process biased to account for hyperedged communities within a plausible microscopic framework, we consider the classification task studied by Zhou *et al.* [46]. We anticipate that the obtained classification outperforms that obtained under the usual random-walk framework, which ignores the annotated hyper structures.

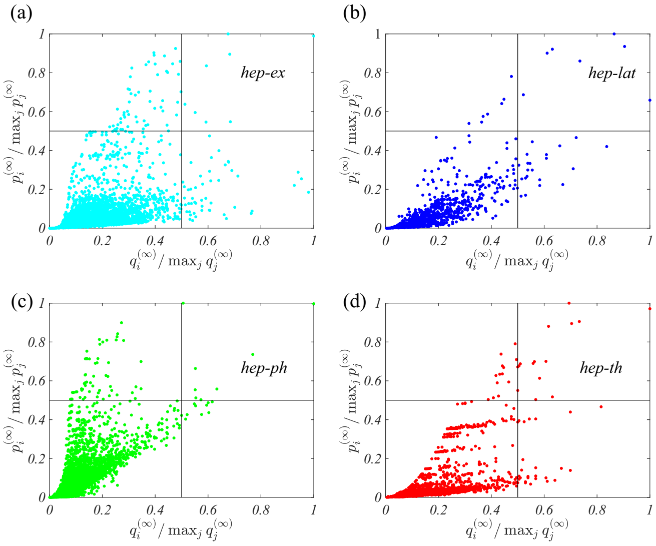

FIG. 7. Publication habits of arXiv-hep. We report the scatter plot of the normalized rankings obtained with the random walk on the projected network $q_i^{(\infty)}$ and the one computed using the RW on hypergraphs $p_i^{(\infty)}$ for the four subdomains of the arXiv-hep domain.

A standard pipeline to analyze a data set starts with the determination of pairwise similarities between the objects to eventually be classified. This implies defining a network that can be studied by means of standard spectral methods. However, similarities often involve groups of objects. In this respect, hypergraphs define the ideal mathematical platform to account for the inherent complexity of the classification problem. More precisely, one can make use of spectral methods based on the hypergraph Laplace matrix to eventually obtain a classification which effectively accounts for high-order interaction as displayed in the data [57].

Following [46], we consider an ensemble of gilled mushrooms from the Agaricus and Lepiota genera and we aim at classifying them into two classes, definitely edible and definitely poisonous (which indeed contains also unknown edibility and not recommended), given their description. Specifically, we used the mushroom database taken from the UCI Machine Learning Depository [58], containing 8124 mushrooms, each one endowed with 22 features, such as cap shape, cap color, odor, and so on (see Appendix F for a complete description of the data set). Here nodes are mushrooms and hyperedges features; we will show that the presence of high-order interactions among features allow one to obtain a very satisfying embedding using only two or three dimensions, a result which is better that the one reported by Zhou *et al.* [46] for an *ad hoc* choice of the free weight parameters, i.e., unitary ones. To this end we build a hypergraph using the above recipe; we compute its random-walk Laplacian and eventually its ensuing spectrum. We list the eigenvalues in ascending order and rename accordingly the eigenvectors. We use the first left eigenvectors,[2] associated with the smallest

[2] In principle, also the right eigenvectors can be used for classification purposes.

TABLE I. Some figures for the arXiv subdomains. The first column shows the subdomain of the arXiv server, while the second one stands for the period of time for which we have extracted the information. Columns 3, 4, and 5 display, respectively, the number of nodes, the number of maximal unique hyperedges, and the number of links in the largest connected component, while in parentheses we show the same values for the whole hypergraph or network. In column 6 we report the size of the largest hyperedge and in the seventh column the maximum degree.

arXiv	Period	Nodes	Hyperedges	Links	max $\|E_\alpha\|$	max k_i
astro-ph	1992–2018	185579 (195729)	136918 (201270)	4602315 (4617912)	81	2732
cond-mat	1992–2018	221415 (243749)	141611 (207939)	1520895 (1551863)	63	1426
cs	1993–2018	136146 (187689)	84184 (139334)	534462 (607560)	65	406
econ	2017–2018	113 (1147)	63 (612)	214 (1295)	5	36
gr-qc	1992–2018	32088 (40316)	25321 (45378)	216355 (228811)	80	511
hep-ex	1992–2018	48460 (55634)	12310 (23249)	1418268 (1435372)	83	1228
hep-lat	1992–2018	10275 (12483)	7439 (14143)	85194 (87254)	72	346
hep-ph	1992–2018	62885 (70324)	50403 (86150)	814746 (823705)	74	1244
hep-th	1992–2018	41814 (51045)	42410 (74136)	144710 (154737)	57	206
math	1992–2018	112203 (159595)	106583 (194312)	279891 (313402)	60	336
nlin	1993–2018	19491 (30445)	12428 (23503)	52089 (64890)	46	312
physics	1996–2018	188142 (240866)	68805 (116611)	1859156 (1950143)	80	891
q-bio	2003–2018	23630 (45103)	9926 (21191)	93127 (142136)	54	176
q-fin	2008–2018	3136 (8721)	2155 (6042)	6851 (13078)	11	66
stat	2008–2018	39422 (57955)	23377 (39366)	130665 (158435)	65	228

eigenvalues, as coordinates of a Euclidean space where to embed the data (see Fig. 8). We observe that since we use a random-walk Laplacian, the first eigenvector, i.e., the one associated with the 0 eigenvalue, is not homogeneous and it contains nontrivial information on the structure of the examined sample. Classes are identified by different colors: Edible mushrooms are reported in blue and poisonous in red. One can visually appreciate that homologous colors cluster in space, hence suggesting that the embedding yields an accurate classification. Indeed, the ground-truth partition of animals into these seven classes and the one obtained by performing a k-means clustering in this three-dimensional space have an adjusted Rand index (ARI) [59] equal to 0.32. For the sake of comparison, the clustering obtained using the method proposed in [46] returns an ARI equal to 0.14, while the one obtained using the eigenvectors of the Laplacian of the projected network have an ARI equal to 0.008. In the latter case the method is less efficient. The classification task is difficult with such a small number of dimensions if the hyperedges are not at play, and the clusters found are not correlated to the ground-truth partition.

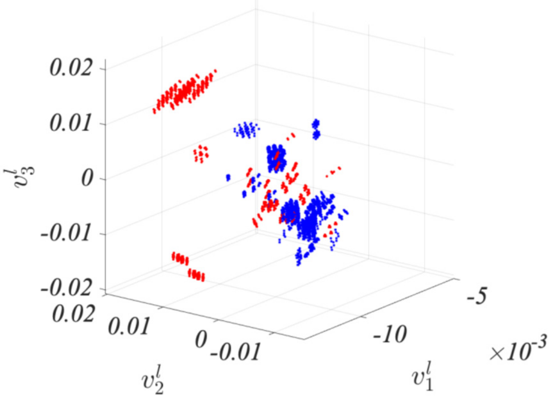

FIG. 8. Classification of the mushrooms according to their features. We report a 3D embedding of the mushroom data set, namely, using the first three eigenvectors. Each combination color refers to a known class, red for poisonous and blue for edible.

We applied our method (and compared it to the one proposed in [46]) to other data sets, ranging from very small to medium sizes, all taken from the UCI Machine Learning Depository [58]: Associate the right contact lenses with a patient (this is a very small data set with only 24 items and 4 features each), determine the animal given distinctive features (this data set contains 101 animals, each one described by 20 features), and evaluate a car (1728 cars and 6 attributes). In all the examined cases the obtained results are more accurate or in line with the one presented in [46] (see Appendix F for a complete description of the data sets).

V. CONCLUSION

Summing up, we have here introduced and studied a class of random walks on hypergraphs which take into account the presence of higher-order interactions. In deriving the transition rates we assumed that the size of the hyperedge linearly correlates with the probability for the walker to perform a jump. In principle, one can relax this assumption and introduce nonlinear transition rates, but exploring further generalizations is left for future investigations. We provided an analytical expression for the ensuing stationary distribution, based on the structural features of the networked system, and compared it to the distribution associated with a traditional random walk performed on the corresponding projected network. More precisely, we proposed a self-consistent recipe grounded in a microscopic physical random process biased by the hyperedges sizes to assign weights to hyperedges. We further characterized the dynamics by comparing the two processes on several synthetic and real-world networks, by means of both numerical simulations and analytical arguments. We showed that our process produces stationary distributions different from those obtained for the corresponding projected network and that prove sensitive to higher-order structure

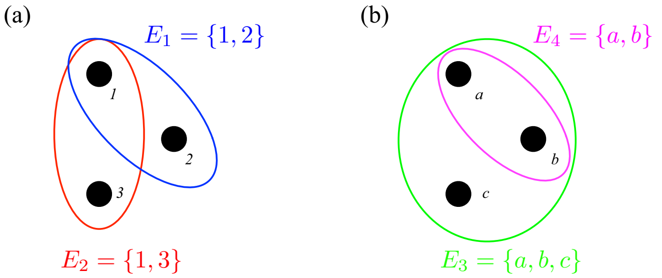

FIG. 9. (a) Simple hypergraph with nodes $V = \{1, 2, 3\}$ and hyperedges $E_1 = \{1, 2\}$ and $E_2 = \{1, 3\}$ and (b) nonsimple hypergraph with nodes $W = \{a, b, c\}$ and hyperedges $E_3 = \{a, b, c\}$ and $E_4 = \{a, b\}$.

in a networked architecture. Our framework was applied to collaboration networks, yielding insights into node ranking and centrality measure, which allow for a richer characterization of individual performances as compared to traditional methods. Moreover, we showed that information embedded in the higher-order walk can be used to achieve accurate classification. In particular, we applied our method to successfully cluster into different families animals with different features, each one representing a hyperedge. The same procedure fails if a simple random walk on the corresponding projected network is considered. Importantly, the proposed Laplacian is equivalent to that stemming from a properly tuned weighted network [53]. Higher-order rankings and refined classifications hence could be immediately obtained by supplying to conventional tools and analysis schemes the weighted adjacency matrix that characterizes the graph with pairwise edges associated with the hypergraph construction. Taken all together, our work sheds light on dynamical processes on networks which are not limited to pairwise interactions and on the complex interplay between the structure and dynamics of higher-order interaction networks. Future applications to machine-learning-based approaches to classification are also envisaged.

ACKNOWLEDGMENTS

We thank an anonymous referee for pointing out an issue with a previous version of Eq. (11). T.C. would like to thank F. Gargiulo for her help in the data collection process. This research used resources of the Plateforme Technologique de Calcul Intensif (PTCI) [60] located at the University of Namur, Belgium, which is supported by the FNRS-FRFC, the Walloon Region, and the University of Namur (Conventions No. 2.5020.11, No. GEQ U.G006.15, No. 1610468, and No. RW/GEQ2016). The PTCI is member of the Consortium des Équipements de Calcul Intensif [61]. F.B. acknowledges partial support from the ERC Synergy Grant No. 810115 (DYNASNET).

APPENDIX A: PROJECTED NETWORK

A hypergraph is simple if each hyperedge does not contain any other hyperedge. We report in Fig. 9 two examples; the hypergraph H_1 with nodes $V = \{1, 2, 3\}$ and hyperedges $E_1 = \{1, 2\}$ and $E_2 = \{1, 3\}$ is simple because either $E_1 \not\subset E_2$ or $E_2 \not\subset E_1$. On the other hand, the hypergraph H_2 with nodes $W = \{a, b, c\}$ and hyperedges $E_3 = \{a, b, c\}$ and $E_4 = \{a, b\}$ is not simple because $E_4 \subset E_3$.

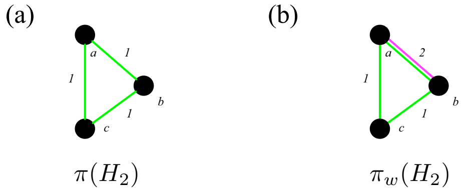

FIG. 10. We propose (a) a standard projection and (b) a weighted projection of the hypergraph H_2 shown in Fig. 9. In the latter case, the edge (a, b) has weight 2 because it belongs to two different hyperedges in H_2.

Once we build the projected network $\pi(H_2)$, starting from the latter hypergraph we get a complete 3-clique, thus losing information on the existence of hyperedge E_4 [see Fig. 10(a)]. Hence, we cannot get back to H_2 by inverting the construction $\pi^{-1}\pi(H_2) \neq H_2$. A possible way to overcome this difficulty is to consider a weighted projection [see Fig. 10(b)] where edges inherit a weight counting the number of different hyperedges they belong to. Observe, however, that for large hyperedge sizes the inversion can be computationally costly because of the combinatorial structure of the problem.

APPENDIX B: TRANSITION PROBABILITY

The aim of this Appendix is to provide some details about the calculation of the generalized transition probabilities which take into account the high-order structure of the hyperedges. To compute the transition probability to jump from i to j, we first count the number of nodes, excluding node i itself, belonging to the same hyperedge of i and j:

$$k_{ij}^H = \sum_\alpha (C_{\alpha\alpha} - 1) e_{i\alpha} e_{j\alpha}, \quad i \neq j, \; k_{ii}^H = 0 \, \forall \, i; \tag{B1}$$

namely, for each hyperedge E_α we consider the number of its nodes minus one, i.e., $C_{\alpha\alpha} - 1$. Then this quantity is added to k_{ij}^H if and only if $e_{i\alpha} = e_{j\alpha} = 1$, that is, if and only if both i and j belong to E_α.

Next we normalize this quantity by considering a uniform choice among the connected hyperedges. Hence, we obtain an initial formula for the transition probability T_{ij} to jump from node i to node j,

$$T_{ij} = \frac{k_{ij}^H}{\sum_{\ell \neq i} k_{i\ell}^H} = \frac{\sum_\alpha (C_{\alpha\alpha} - 1) e_{i\alpha} e_{j\alpha}}{\sum_{\ell \neq i} \sum_\alpha (C_{\alpha\alpha} - 1) e_{i\alpha} e_{\ell\alpha}}, \tag{B2}$$

so that $\sum_j T_{ij} = 1 \, \forall \, i$.

Equation (B2) can be rewritten in an equivalent form, which allows one to draw a comparison with the transition probability for unbiased random walks on networks. Indeed, by recalling the definition of $C_{\alpha\beta} = (e^T e)_{\alpha\beta} = \sum_\ell e_{\alpha\ell}^T e_{\ell\beta} = \sum_\ell e_{\ell\alpha} e_{\ell\beta}$, we get $C_{\alpha\alpha} = \sum_\ell e_{\ell\alpha} e_{\ell\alpha}$ and then

$$\sum_\alpha C_{\alpha\alpha} e_{i\alpha} e_{j\alpha} = \sum_\alpha e_{i\alpha} C_{\alpha\alpha} e_{\alpha j}^T = (e\hat{C}e^T)_{ij}, \tag{B3}$$

where $\hat{C}$ is a diagonal matrix: The diagonal of $\hat{C}$ coincides with that of C and its off-diagonal is identically equal to zero.

This allows us to rewrite Eq. (B1) in a more compact way [Eq. (2)]

$$k_{ij}^H = \sum_\alpha (C_{\alpha\alpha} - 1) e_{i\alpha} e_{j\alpha} = (e\hat{C}e^T)_{ij} - (ee^T)_{ij}$$
$$= (e\hat{C}e^T)_{ij} - A_{ij}\,\forall\, i \neq j,$$

where in the last step we used the definition of the adjacency matrix of the hypergraph. We thus eventually get Eq. (3).

We observe that this equation remains valid even for nonsimple hypergraphs. For instance, using again the hypergraph H_2 shown in Fig. 9, where the hyperedge E_4 is properly included into E_3, we get

$$k_{ab} = (E_3 - 1) + (E_4 - 1) = 2 + 1, \quad k_{ac} = E_3 - 1 = 2$$

and thus the transition probabilities

$$T_{ab} = \tfrac{3}{5}, \quad T_{ac} = \tfrac{2}{5},$$

so the transition from a to b is 1.5 times more probable than to c because a and b share two hyperedges. Among nonsimple hypergraphs, one has to account for the fact that hyperedges are repeated several times. The theory proposed here holds true also for weighted hyperedges.

Nonlinear transition rates

In deriving the transition rates (3), we assumed that the size of the hyperedge linearly correlates with the probability for the walker to perform a jump; one can of course relax this assumption and introduce nonlinear transition rates. In other words, one can add a bias in (B1) in the selection rule for a target node j, as operated by a walker sitting on node i. For example, one can posit

$$k_{ij}^{(H,\gamma)} = \sum_\alpha (C_{\alpha\alpha} - 1)^\gamma e_{i\alpha} e_{j\alpha}, \quad i \neq j, \; k_{ii}^{(H,\gamma)} = 0\,\forall\, i. \tag{B4}$$

In this way, large hyperedges are even more favored if $\gamma > 0$, while the opposite happens if $\gamma < 0$, and we eventually get for the transition probabilities

$$T_{ij}^{(\gamma)} = \frac{k_{ij}^{(H,\gamma)}}{\sum_{\ell\neq i} k_{i\ell}^{(H,\gamma)}}.$$

APPENDIX C: STATIONARY SOLUTION

Given the transition probability stored in the matrix $\mathbf{T}$, we can obtain the analytical solution for the stationary state $\mathbf{p}^{(\infty)}$ defined by $\mathbf{p}^{(\infty)} = \mathbf{p}^{(\infty)}\mathbf{T}$. By recalling Eq. (9), we can straightforwardly verify that it solves the fixed point equation for the governing dynamics. To this end one needs to plug Eq. (9) into Eq. (8) and recall the definition (3) for $T_{ji} = k_{ji}^H / d_j^H$,

$$\sum_j \left(\frac{d_j^H}{\sum_m d_m^H}\right)\left(\frac{k_{ji}^H}{d_j^H} - \delta_{ji}\right) = \sum_j \frac{k_{ji}^H}{\sum_m d_m^H} - \frac{d_i^H}{\sum_m d_m^H} = 0, \tag{C1}$$

where the last step has been obtained by recalling that $k_{jj}^H = 0$ and thus $\sum_j k_{ji}^H = d_i^H$.

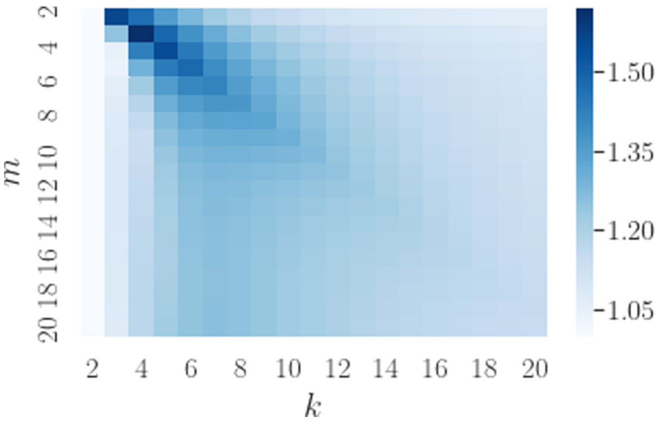

FIG. 11. Star-clique model: ratio between the Gini coefficient of the stationary state on the hypergraph and the projected network, varying the size of the clique k and of the star m.

The analytical formula for the stationary distribution allows us to better understand the ranking emerging from the random walk on the hypergraph and in particular the inversion phenomenon as discussed in the main text (a top-ranked node for the high-order structure loses its leading position when studied in the projected network, or vice versa).

APPENDIX D: HETEROGENEITY OF THE STATIONARY SOLUTION

The stationary solution that we obtain from a random walk on a hypergraph is very different from the one we can get from the corresponding projected network, the former being more sensitive to the organization in groups. The heterogeneity of the state, i.e., the difference among the occupation probabilities of the different nodes at equilibrium, can be quantified by making use of the Gini coefficient.

Figure 11 reports on the ratio between the coefficient G computed for the hypergraph and for the projected network of Fig. 2, at varying m, the size of the star, and k, the size of the clique. In contrast, Fig. 12 shows the heterogeneity for the model which goes from a 1D lattice to a fully connected network, by subsequently adding the links (see Fig. 5). The red points show the Gini coefficient for the hypergraph, while the green ones are plotted for the projected network, at varying ℓ, the number of links in the graph.

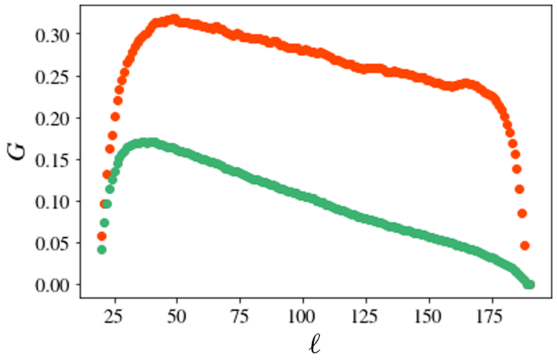

FIG. 12. Lattice to fully connected model: Gini coefficient for the stationary state of the random walk on the hypergraph (red) and on the corresponding projected network (green).

From the results presented in these figures one can appreciate that the Gini coefficient associated with the stationary solution for the random walk on the hypergraph is always larger than the same quantity computed for the random walk on the projected network. This implies thus that the distribution of walkers on the hypergraph is more heterogeneous than for the projected network.

APPENDIX E: COAUTHORSHIP NETWORKS FROM arXiv

The collaboration network is one of the most representative examples of hypergraph; nodes are authors and hyperedges are groups of authors that collaborated to accomplish a task, e.g., write a scientific paper. For this reason we decided to apply the method that we developed to the coauthorship networks extracted from the online preprints platform arXiv and hence analyze the nodes ranking obtained using the two processes.

In this Appendix we report some results for the coauthorship hypergraph for the subdomains of arXiv since their existence up to 2018 included (second column Table I). In each subdomain, we gathered all the papers and then extracted the authors names, so creating a hyperedge whose nodes are the authors. We thus obtain a set of nodes $V^{(1)}$ and hyperedges $E^{(1)}$ and also the edges of the associated projected network $E_q^{(1)}$. Such quantities are reported in parentheses in the third, fourth, and fifth columns of the Table I. Once the hypergraph has been built, we identify the largest connected component that will contain the nodes $V^{(cc)}$; then we identify all the maximal hyperedges, i.e., not properly contained in any other larger hyperedge, and unique hyperedges $E^{(cc)}$ and the edges of the associated projected network $E_q^{(cc)}$. Columns 3–5 of the table show such values. Finally, we compute the largest hyperedge and the largest node degree in the maximal connected component (columns 6 and 7). For instance, in the arXiv-cs there is a node that belongs to a hyperedge of size 65 and that is linked to 406 other nodes; this means that this researcher has signed a paper with 64 other researchers and in total had 406 different collaborators with whom the researcher has written a paper. Let us also observe that because of the maximality and uniqueness assumptions, we do not know if the researcher has coauthored other papers with a subset of the 64 scholars. Moreover, because we used unweighted networks, we also cannot estimate how many papers the researcher wrote with the 406 collaborators. Let us recall that the need for the maximality and uniqueness is only to compare the results with the projected network, while our method works also without these assumptions.

Authors and articles in each subdomain follow different "rules" and "habits" of publication and writing papers. However, the distributions of node degrees, i.e., number of different collaborators per author, and of hyperedges size, i.e., number of coauthors in papers, exhibit quite similar shapes across the domains, e.g., broad tails (see Fig. 13 for the degree distribution and Fig. 14 for the hyperedges size distribution).

As already stated, the random walk on the hypergraph gives more relevance to the size of the hyperedge, i.e., the number of coauthors, while the same process on a network emphasizes the number of different collaborators. Let us recall that here we considered unweighted hypergraphs and networks. We can thus use these approaches to distinguish the different publication habits in the considered subdomains. To this aim we first normalize the stationary probabilities $p_i^{(\infty)}$ for the hypergraph and $q_i^{(\infty)}$ for the projected network, with respect to their maximum value, to be able to compare sets containing different amounts of data, and then we report in the plane with coordinates $(q_i^{(\infty)}/\max_j q_j^{(\infty)}, p_i^{(\infty)}/\max_j p_j^{(\infty)})$ the scatter plot of the data (each point is an author in the maximal connected component of the hypergraph), separated into different subdomains (see Fig. 15).

If the computed rankings were (almost) the same, the data would (almost) lie on the main diagonal; deviation from this results in novel information conveyed by the random walk on the hypergraph. Besides the region delimited by $q_i^{(\infty)}/\max_j q_j^{(\infty)} \leqslant 1/2$ and $p_i^{(\infty)}/\max_j p_j^{(\infty)} \leqslant 1/2$, associated with authors having written a few articles (low degree) and in small groups, we identify three interesting zones associated (roughly speaking) with the squares: $[1/2, 1] \times [0, 1/2]$ (bottom right), $[0, 1/2] \times [1/2, 1]$ (top left), and $[1/2, 1] \times [1/2, 1]$ (top right). Authors in the top right square are top ranked in both processes; they have hence written a large number of papers with different collaborators, i.e., large degree, but also they have participated in a relevant number of papers with many coauthors, i.e., large hyperedge size. Scholars in the bottom right square are better ranked by the random walk on the network. This means that they have written several papers but with a small number of coauthors (see, e.g., the panel "physics" in Fig. 15). Finally, scholars in the top left square behave in the opposite way: They have participated in a small number of papers but written by many authors (see, e.g., panels gr-qc, q-bio, and stat in Fig. 15).

APPENDIX F: THE UCI DATABASES

We gathered several databases from the UCI Machine Learning Depository [58] involving multivariate, categorical, and numerical variables. From each database we build a hypergraph where nodes are the items and the hyperedges represent their features. The data sets are manually annotated and so the ground truth is available; they have been created to provide a benchmark for machine learning tools, to test their capacity to correctly assign each item to the right class based on the associated features. To proceed in the analysis and make the database uniform, we transform the categorical variables into Boolean ones, for instance, in the case of animals the hair feature becomes a 0-1 variable. Moreover, to have hyperedges containing only 1's and 0's, we process some of the available input; for instance, again in the case of animals, the feature associated with the number of legs, i.e., 0, 2, 4, 6, and 8, gives rise to five new Boolean features, i.e., has 0 legs, has 2 legs, has 4 legs, has 6 legs, and has 8 legs.

Items are the nodes of the hypergraph and the features are the hyperedges; hence all items sharing the same feature belong to the same hyperedge. The projected network is obtained by making a complete clique from each hyperedge; that is to create a link between all the nodes sharing the same property. We observe that this can also be seen as the projection of the bipartite network where there are two kinds of nodes, items and features, each one linked only to nodes of the other kind. We thus compute the spectrum of the hypergraph

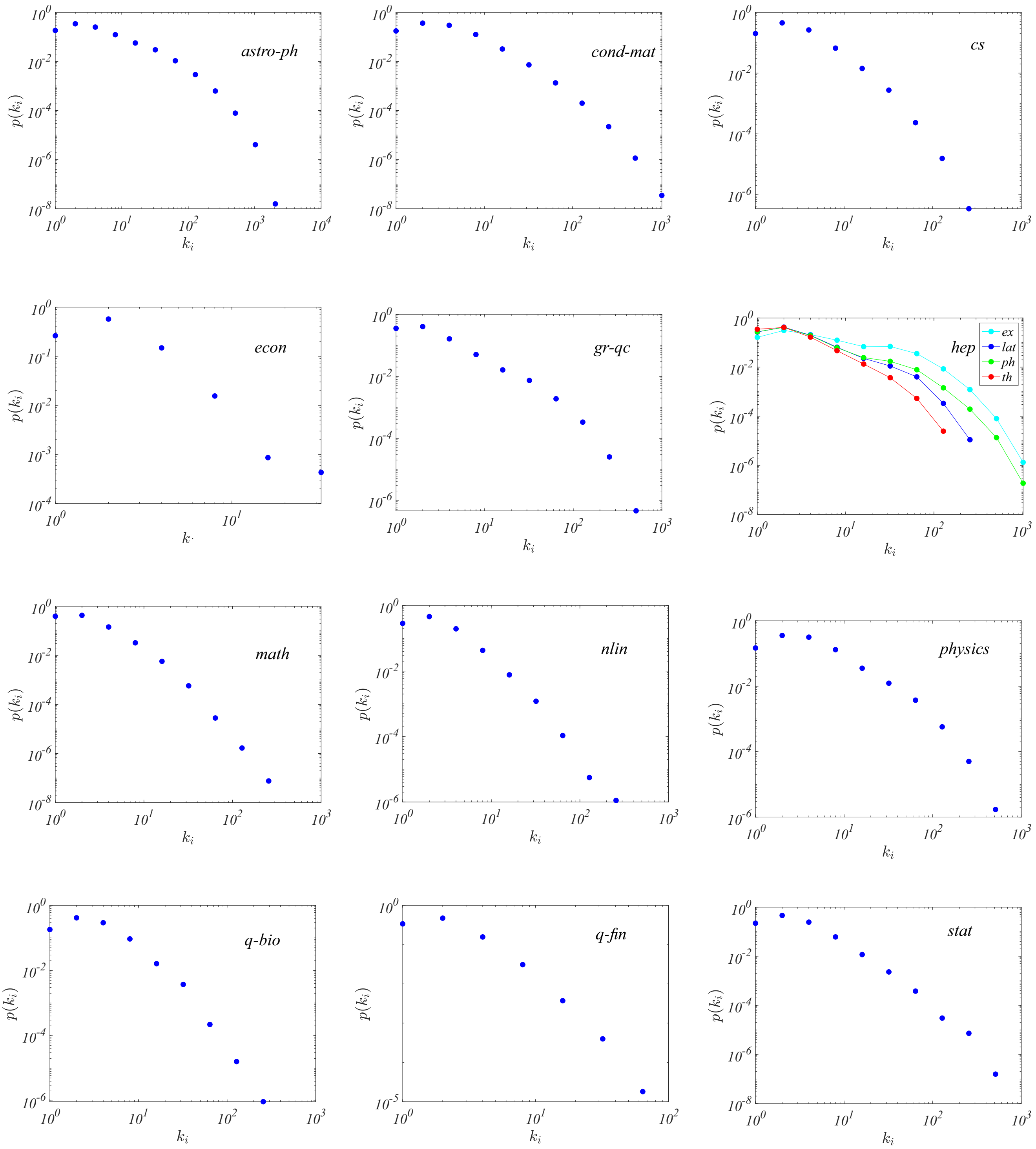

FIG. 13. Degree distribution. We report for the arXiv subdomains the probability distribution of node degrees $p(k_i)$ associated with the maximal connected component. In all cases, we observe a broad distribution: Notice that the arXiv-econ has a relatively small number of papers and authors because of its young age (2017–2018) and thus also that the maximal degree, i.e., number of papers written by an author, is quite small.

Laplacian, the one proposed by Zhou *et al.* [46] with unitary weights, the Bolla Laplacian [62], the Rodríguez Laplacian [63], and the one for the projected network. In each case, we rank eigenvalues in ascending order, 0 being the smallest one. We then accordingly rename the associated left eigenvector and use the first few to embed the data set in low-dimensional Euclidean space. The results (see Table II) are presented using quantitative ARI scores comparing the classification resulting from $k = 2$ and $k = 3$ embedding with the ground truth; we can observe that the classification based on our Laplacian performs very well, exhibiting the largest ARI in all cases but one (the zoo data set using a 3D embedding where the method

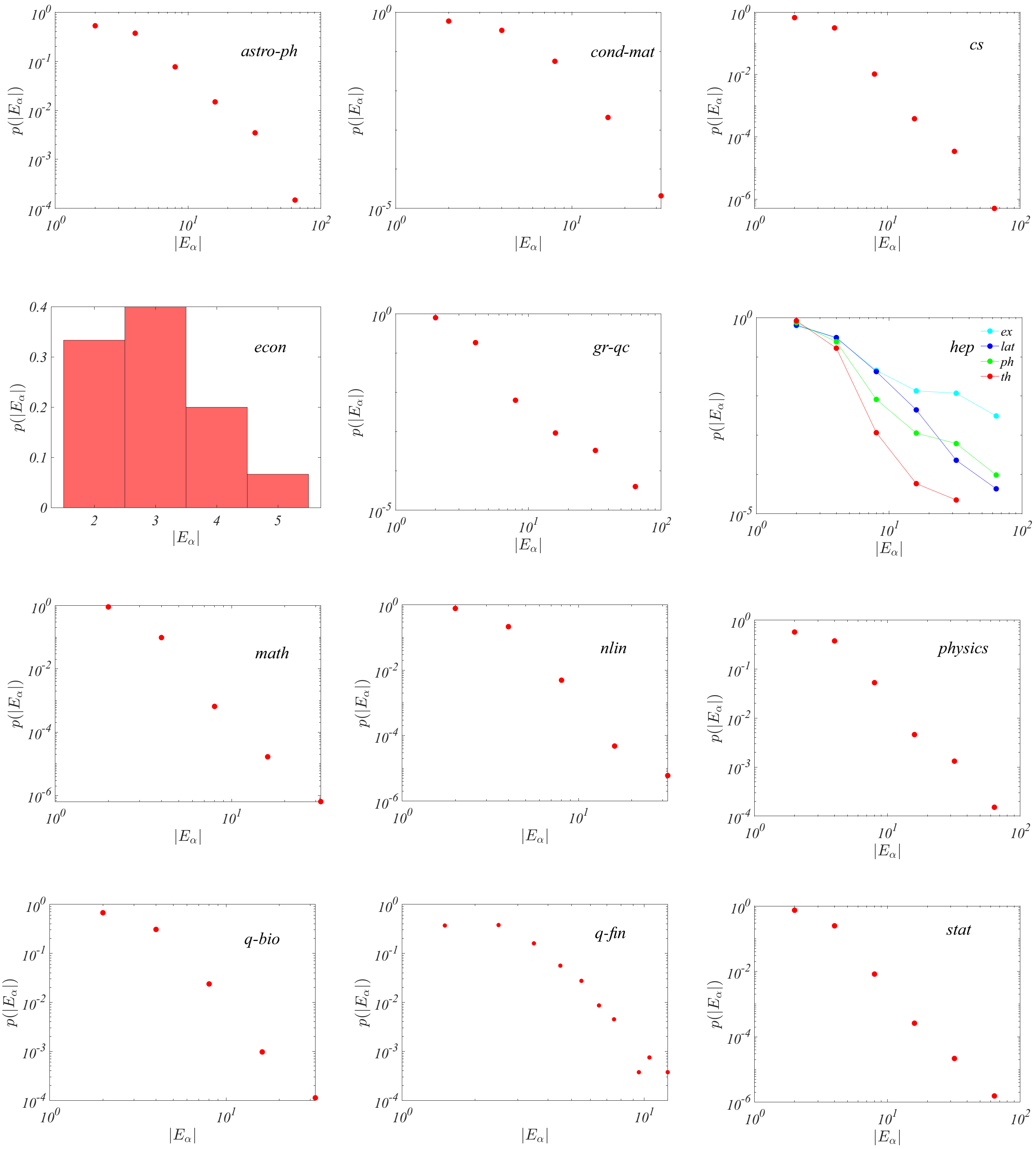

FIG. 14. Hyperedges size distribution. We report for the arXiv subdomains the probability distribution of hyperedges size $p(|E_\alpha|)$ associated with the maximal connected component. In all the cases we observe a broad distribution, except for the arXiv-econ, for which the number of papers and authors is relatively small because of its young age (2017–2018) and thus also the maximal hyperedge size, i.e., number of coauthors of a paper, is quite small. For this reason we report data in the form of a histogram.

by Zhou *et al.* works better) and in particular with large data sets. We do not report the ARI values for the projected networks because they are very low. Indeed, classification performances based on the projected unweighted network are significantly worse as those obtained when preserving the high-order information.

1. Lenses database

The database contains several features of patients with poor eyesight and aims to associating each one with the appropriately chosen contact lens: hard contact lenses, soft contact lenses, or should not be fitted with contact lenses.

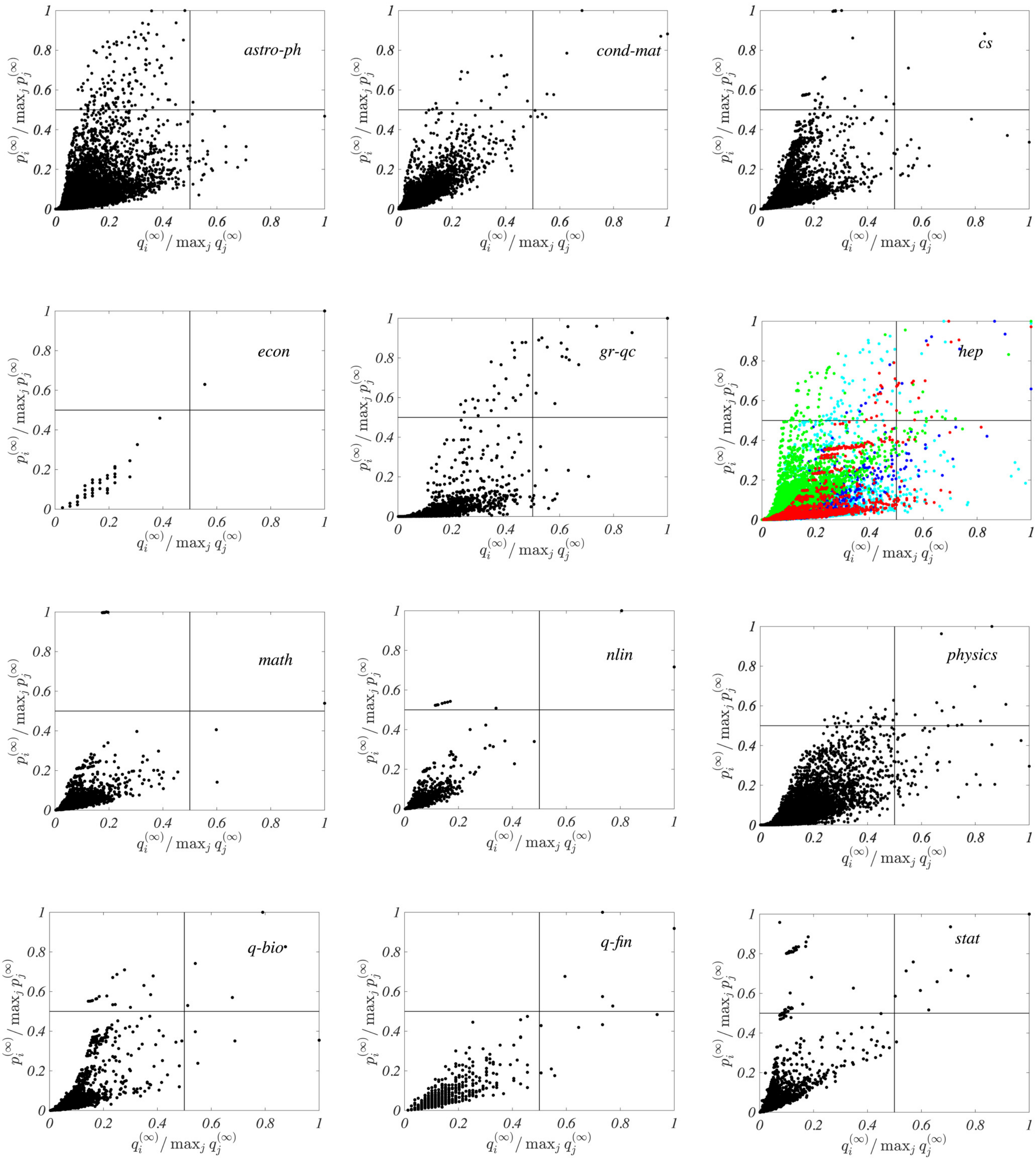

FIG. 15. Comparison of the rankings in the arXiv community. We report the scatter plot of the normalized rankings obtained with the random walk on network $q_i^{(\infty)}$ and the one computed using the random walk on hypergraphs $p_i^{(\infty)}$.

There are thus three classes. The following are the patients' features.

(i) Age of the patient: (1) young, (2) prepresbyopic, or (3) presbyopic.

(ii) Spectacle prescription: (1) myope or (2) hypermetrope.

(iii) Astigmatic: (1) no or (2) yes.

(iv) Tear production rate: (1) reduced or (2) normal.

All the features but the first one are already in a Boolean format. We thus introduced three additional features.

(v) Is the patient young?: (1) yes or (2) no.

TABLE II. The ARI coefficients. We report the ARI coefficients for $k = 2$ and $k = 3$ embedding, computed for several databases (first column) using different models, our version of the random-walk Laplacian on hypergraph (fourth and fifth columns), the Zhou *et al.* Laplacian [46] with unitary weights (sixth and seventh columns), the Bolla Laplacian [62] (eighth and ninth columns), and the Rodríguez Laplacian [63] (tenth and eleventh columns). The second and third columns show some figures of the databases. The figures emphasized in boldface denote the largest values of the ARI index for the different databases.

			ARI $\mathbf{L}$		ARI $\mathbf{L}^z$		ARI $\mathbf{L}^b$		ARI $\mathbf{L}^r$	
Database	Number of items	Number of features	$k = 2$	$k = 3$	$k = 2$	$k = 3$	$k = 2$	$k = 3$	$k = 2$	$k = 3$
lenses	24	9	**0.1427**	0.1285	0.0323	−0.0357	−0.0221	0.0749	0.0370	0.1190
zoo	101	20	0.4350	0.5216	0.5074	**0.5668**	0.4647	0.5332	0.2650	0.5495
car evaluation	1728	6	0.0525	**0.0552**	0.0163	0.0075	0.0001	0.0108	0.0082	0.0143
mushroom	8124	105	**0.6032**	0.2852	0.1587	0.1314	0.1508	0.1406	0.1836×10^{-4}	0.1836×10^{-4}

(vi) Is the patient prepresbyopic?: (1) yes or (2) no.

(vii) Is the patient presbyopic?: (1) yes or (2) no.

In this way the total number of features is 9 and the number of items is 24.

2. Zoo database

The zoo data set contains 101 animals, each one endowed with 15 Boolean features, whose value is thus yes or not, e.g., hair, feathers, eggs, milk, airborne, aquatic, predator, toothed, backbone, breathes, venomous, fins, tail, domestic, and cat size. There is also a further class that reports on the number of legs, i.e., 0, 2, 4, 6, and 8. To homogenize the data set we decided to introduce five new Boolean features to replace the last one, the new ones being has 0 legs, has 2 legs, has 4 legs, has 6 legs, and has 8 legs. The data set is manually annotated; hence for each animal we have the right class it belongs to, e.g. mammal, bird, reptile, fish, amphibian, bug, and invertebrate. In conclusion, we have 20 features and 7 classes.

3. Car evaluation database

The database contains 1728 cars, each one characterized by 6 attributes.

(i) Buying price: very high, high, medium, or low.

(ii) Maintenance price: very high, high, medium, or low.

(iii) Number of doors: 2, 3, 4, 5, or more.

(iv) Persons to carry: 2, 4, or more.

(v) Size of luggage boot: small, medium, or big.

(vi) Estimated safety of the car: low, medium, or high.

The goal is to decide if a car is unacceptable, acceptable, good, or very good. There are thus 4 classes and 21 features, once we transform the previous 6 into Boolean ones.

4. Mushroom database

The data set contains 8124 gilled mushrooms in the Agaricus and Lepiota genera and each specimen is identified as definitely edible, definitely poisonous, or of unknown edibility and not recommended. This latter class was combined with the poisonous one. There are thus 2 classes. Each mushroom has the following 21 attributes.

(i) Cap shape: bell, conical, convex, flat, knobbed, or sunken.

(ii) Cap surface: fibrous, grooves, scaly, or smooth.

(iii) Cap color: brown, buff, cinnamon, gray, green, pink, purple, red, white, or yellow.

(iv) Bruises: bruises or no bruises.

(v) Odor: almond, anise, creosote, fishy, foul, musty, none, pungent, or spicy.

(vi) Gill attachment: attached, descending, free, or notched.

(vii) Gill spacing: close, crowded, or distant.

(viii) Gill size: broad or narrow.

(ix) Gill color: black, brown, buff, chocolate, gray, green, orange, pink, purple, red, white, or yellow.

(x) Stalk shape: enlarging or tapering.

(xi) Stalk surface above the ring: fibrous, scaly, silky, or smooth.

(xii) Stalk surface below the ring: fibrous, scaly, silky, or smooth.

(xiii) Stalk color above the ring: brown, buff, cinnamon, gray, orange, pink, red, white, or yellow.

(xiv) Stalk color below the ring: brown, buff, cinnamon, gray, orange, pink, red, white, or yellow.

(xv) Veil type: partial or universal.

(xvi) Veil color: brown, orange, white, or yellow.

(xvii) Ring number: none, one, or two.

(xviii) Ring type: cobwebby, evanescent, flaring, large, none, pendant, sheathing, or zone.

(ixx) Spore print color: black, brown, buff, chocolate, green, orange, purple, white, or yellow.

(xx) Population: abundant, clustered, numerous, scattered, several, or solitary.

(xxi) Habitat: grasses, leaves, meadows, paths, urban, waste, or woods.

Eventually, we end up with 105 features once they all are transformed into Boolean ones.

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

[2] A.-L. Barabási, *Network Science* (Cambridge University Press, Cambridge, 2016).

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

[4] R. Albert and A.-L. Barabási, Statistical mechanics of complex networks, Rev. Mod. Phys. **74**, 47 (2002).

[5] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, Complex networks: Structure and dynamics, Phys. Rep. **424**, 175 (2006).

[6] C. Castellano, S. Fortunato, and V. Loreto, Statistical physics of social dynamics, Rev. Mod. Phys. **81**, 591 (2009).

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

[8] R. Lambiotte, M. Rosvall, and I. Scholtes, From networks to optimal higher-order models of complex systems, Nat. Phys. **15**, 313 (2019).

[9] I. Scholtes, N. Wider, R. Pfitzner, A. Garas, C. J. Tessone, and F. Schweitzer, Causality-driven slow-down and speed-up of diffusion in non-Markovian temporal networks, Nat. Commun. **5**, 5024 (2014).

[10] M. Rosvall, A. V. Esquivel, A. Lancichinetti, J. D. West, and R. Lambiotte, Memory in network flows and its effects on spreading dynamics and community detection, Nat. Commun. **5**, 4630 (2014).

[11] M. De Domenico, C. Granell, M. A. Porter, and A. Arenas, The physics of spreading processes in multilayer networks, Nat. Phys. **12**, 901 (2016).

[12] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, Multilayer networks, J. Complex Networks **2**, 203 (2014).

[13] F. Battiston, V. Nicosia, and V. Latora, The new challenges of multiplex networks: Measures and models, Eur. Phys. J.: Spec. Top. **226**, 401 (2017).

[14] A. R. Benson, D. F. Gleich, and J. Leskovec, Higher-order organization of complex networks, Science **353**, 163 (2016).

[15] A. R. Benson, R. Abebe, M. T. Schaub, A. Jadbabaie, and J. Kleinberg, Simplicial closure and higher-order link prediction, Proc. Natl. Acad. Sci. USA **115**, E11221 (2018).

[16] I. Iacopini, G. Petri, A. Barrat, and V. Latora, Simplicial models of social contagion, Nat. Commun. **10**, 2485 (2019).

[17] J. Grilli, G. Barabás, M. J. Michalska-Smith, and S. Allesina, Higher-order interactions stabilize dynamics in competitive network models, Nature (London) **548**, 210 (2017).

[18] K. Devriendt and P. Van Mieghem, The simplex geometry of graphs, J. Complex Networks **7**, 469 (2019).

[19] O. T. Courtney and G. Bianconi, Generalized network structures: The configuration model and the canonical ensemble of simplicial complexes, Phys. Rev. E **93**, 062311 (2016).

[20] G. Petri and A. Barrat, Simplicial Activity Driven Model, Phys. Rev. Lett. **121**, 228301 (2018).

[21] C. Berge, *Graphs and Hypergraphs*, North-Holland Mathematical Library Vol. 6 (American Elsevier, New York, 1973).

[22] E. Estrada and J. A. Rodríguez-Velázquez, Subgraph centrality and clustering in complex hyper-networks, Physica A **364**, 581 (2006).

[23] G. Ghoshal, V. Zlatić, G. Caldarelli, and M. E. J. Newman, Random hypergraphs and their applications, Phys. Rev. E **79**, 066118 (2009).

[24] G. Petri, P. Expert, F. Turkheimer, R. Carhart-Harris, D. Nutt, P. J. Hellyer, and F. Vaccarino, Homological scaffolds of brain functional networks, J. R. Soc. Interface **11**, 20140873 (2014).

[25] A. Patania, G. Petri, and F. Vaccarino, The shape of collaborations, EPJ Data Sci. **6**, 18 (2017).

[26] R. M. May, Will a large complex system be stable? Nature (London) **238**, 413 (1972).

[27] S. Allesina and S. Tang, Stability criteria for complex ecosystems, Nature (London) **483**, 205 (2012).

[28] L. M. Pecora and T. L. Carroll, Master Stability Functions for Synchronized Coupled Systems, Phys. Rev. Lett. **80**, 2109 (1998).

[29] S. Redner, *A Guide to First-Passage Processes* (Cambridge University Press, Cambridge, 2001).

[30] J. D. Noh and H. Rieger, Random Walks on Complex Networks, Phys. Rev. Lett. **92**, 118701 (2004).

[31] M. E. J. Newman, A measure of betweenness centrality based on random walks, Soc. Networks **27**, 39 (2005).

[32] M. Rosvall and C. T. Bergstrom, Maps of random walks on complex networks reveal community structure, Proc. Natl. Acad. Sci. USA **105**, 1118 (2008).

[33] V. Nicosia, M. De Domenico, and V. Latora, Characteristic exponents of complex networks, Europhys. Lett. **106**, 58005 (2014).

[34] J. Gómez-Gardeñes and V. Latora, Entropy rate of diffusion processes on complex networks, Phys. Rev. E **78**, 065102(R) (2008).

[35] G. Cencetti, F. Battiston, D. Fanelli, and V. Latora, Reactive random walkers on complex networks, Phys. Rev. E **98**, 052302 (2018).

[36] P. S. Skardal and S. Adhikari, Dynamics of nonlinear random walks on complex networks, J. Nonlinear Sci. **29**, 1419 (2019).

[37] M. Asllani, T. Carletti, F. Di Patti, D. Fanelli, and F. Piazza, Hopping in the Crowd to Unveil Network Topology, Phys. Rev. Lett. **120**, 158301 (2018).

[38] M. Starnini, A. Baronchelli, A. Barrat, and R. Pastor-Satorras, Random walks on temporal networks, Phys. Rev. E **85**, 056115 (2012).

[39] J. Petit, M. Gueuning, T. Carletti, B. Lauwens, and R. Lambiotte, Random walk on temporal networks with lasting edges, Phys. Rev. E **98**, 052307 (2018).

[40] J. Petit, R. Lambiotte, and T. Carletti, Classes of random walks on temporal networks with competing timescales, Appl. Netw. Sci. **4**, 72 (2019).

[41] M. De Domenico, A. Solé-Ribalta, S. Gómez, and A. Arenas, Navigability of interconnected networks under random failures, Proc. Natl. Acad. Sci. USA **111**, 8351 (2014).

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

[43] M. T. Schaub, A. R. Benson, P. Horn, G. Lippner, and A. Jadbabaie, Random walks on simplicial complexes and the normalized Hodge Laplacian, arXiv:1807.05044 [SIAM (to be published)].

[44] L. Lu and X. Peng, in *International Workshop on Algorithms and Models for the Web-Graph*, edited by A. Frieze, P. Horn, and P. Prałat, Lecture Notes in Computer Science Vol. 6732 (Springer, Berlin, 2011), pp. 14–25.

[45] A. Helali and M. Löwe, Hitting times, commute times, and cover times for random walks on random hypergraphs, Stat. Probab. Lett. **154**, 108535 (2019).

[46] D. Zhou, J. Huang, and B. Schölkopf, in *Proceedings of the Conference on Advances in Neural Information Processing Systems 2006*, edited by B. Schölkopf, J. C. Platt, and T. Hoffman (MIT Press, Cambridge, 2007), pp. 1601–1608.

[47] J. T. Matamalas, S. Gómez, and A. Arenas, Abrupt phase transition of epidemic spreading in simplicial complexes, arXiv:1910.03069 [Phys. Rev. Res. (to be published)].

[48] B. Jhun, M. Jo, and B. Kahng, Simplicial SIS model in scale-free uniform hypergraph, J. Stat. Mech. (2019) 123207.

[49] L. Neuhäuser, A. Mellor, and R. Lambiotte, Multi-body interactions and non-linear consensus dynamics on networked systems, arXiv:1910.09226.

[50] G. F. de Arruda, G. Petri, and Y. Moreno, Social contagion models on hypergraphs, arXiv:1909.11154.

[51] M. Del Vicario, G. Vivaldo, A. Bessi, F. Zollo, A. Scala, G. Caldarelli, and W. Quattrociocchi, Echo chambers: Emotional contagion and group polarization on Facebook, Sci. Rep. **6**, 37825 (2016).

[52] M. Belkin and P. Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation, Neural Comput. **15**, 1373 (2002).

[53] U. Chitra and B. J. Raphael, Random walks on hypergraphs with edge-dependent vertex weights, arXiv:1905.08287.

[54] S. Brin and L. Page, The anatomy of a large-scale hypertextual Web search engine, Comput. Networks ISDN Syst. **30**, 107 (1998).

[55] L. Page, S. Brin, R. Motwani, and T. Winograd, The pagerank citation ranking: Bringing order to the web, Technical Report No. 1999-66, Stanford InfoLab (1999).

[56] F. Gargiulo, A. Caen, R. Lambiotte, and T. Carletti, The classical origin of modern mathematics, EPJ Data Sci. **5**, 26 (2016).

[57] L. H. Tran, L. H. Tran, H. Trang, and L. T. Hieu, Combinatorial and random walk hypergraph Laplacian eigenmaps, Int. J. Mach. Learn. Comput. **5**, 462 (2015).

[58] D. Dua and C. Graff, UCI Machine Learning Repository, 2017, available at http://archive.ics.uci.edu/ml.

[59] L. Hubert and P. Arabie, Comparing partitions, J. Classif. **2**, 193 (1985).

[60] http://www.ptci.unamur.be.

[61] http://www.ceci-hpc.be.

[62] M. Bolla, Spectra, Euclidean representations and clusterings of hypergraphs, Discrete Math. **117**, 19 (1993).

[63] J. A. Rodríguez, On the Laplacian eigenvalues and metric parameters of hypergraphs, Linear Multilinear Algebra **50**, 1 (2002).