

## Vector centrality in hypergraphs

K. Kovalenko<sup>a,1</sup>, M. Romance<sup>b,n,o,1</sup>, E. Vasilyeva<sup>a,c,1</sup>, D. Aleja<sup>b,n,o</sup>, R. Criado<sup>b,n,o</sup>, D. Musatov<sup>a,d,e</sup>, A.M. Raigorodskii<sup>a,d,f,g</sup>, J. Flores<sup>b</sup>, I. Samoylenko<sup>h</sup>, K. Alfaro-Bittner<sup>b,n,\*</sup>, M. Perc<sup>i,j,k,l</sup>, S. Boccaletti<sup>a,b,m,n</sup>

<sup>a</sup> Moscow Institute of Physics and Technology, 9 Institutskiy per., Dolgoprudny 141701, Russia

<sup>b</sup> Departamento M.A.C.I.M.T.E, Universidad Rey Juan Carlos, Calle Tulipán s/n, 28933 Móstoles, Madrid, Spain

<sup>c</sup> P.N. Lebedev Physical Institute of the Russian Academy of Sciences, 53 Leninsky prosp., 119991 Moscow, Russia

<sup>d</sup> Caucasus Mathematical Center at Adyghe State University, ul. Pervomaiskaya, 208, Maykop 385000, Russia

<sup>e</sup> Russian Academy of National Economy and Public Administration, pr. Vernadskogo, 84, Moscow 119606, Russia

<sup>f</sup> Mechanics and Mathematics Faculty, Moscow State University, Leninskie Gory, 1, Moscow 119991, Russia

<sup>g</sup> Institute of Mathematics and Computer Science, Buryat State University, ul. Ranzhurova, 5, Ulan-Ude 670000, Russia

<sup>h</sup> National Research University Higher School of Economics, 6 Usacheva str., Moscow 119048, Russia

<sup>i</sup> Faculty of Natural Sciences and Mathematics, University of Maribor, Koroška cesta 160, 2000 Maribor, Slovenia

<sup>j</sup> Department of Medical Research, China Medical University Hospital, China Medical University, Taichung 404332, Taiwan

<sup>k</sup> Complexity Science Hub Vienna, Josefstadtstraße 39, 1080, Vienna, Austria

<sup>l</sup> Alma Mater Europaea, Slovenska ulica 17, 2000 Maribor, Slovenia

<sup>m</sup> CNR - Institute of Complex Systems, Via Madonna del Piano 10, I-50019 Sesto Fiorentino, Italy

<sup>n</sup> Laboratory of Mathematical Computation on Complex Networks and their Applications, Universidad Rey Juan Carlos, Calle Tulipán s/n, Móstoles, 28933 Madrid, Spain

<sup>o</sup> Data, Complex Networks, and Cybersecurity Research Institute, Universidad Rey Juan Carlos, 28028, Madrid, Spain

### ARTICLE INFO

#### Article history:

Received 30 May 2022

Accepted 27 June 2022

Available online xxxx

#### Keywords:

Centrality

Hypergraphs

Networks

### ABSTRACT

Identifying the most influential nodes in networked systems is of vital importance to optimize their function and control. Several scalar metrics have been proposed to that effect, but the recent shift in focus towards network structures which go beyond a simple collection of dyadic interactions has rendered them void of performance guarantees. We here introduce a new measure of node's centrality, which is no longer a scalar value, but a vector with dimension one lower than the highest order of interaction in a hypergraph. Such a vectorial measure is linked to the eigenvector centrality for networks containing only dyadic interactions, but it has a significant added value in all other situations where interactions occur at higher-orders. In particular, it is able to unveil different roles which may be played by the same node at different orders of interactions – information that is otherwise impossible to retrieve by single scalar measures. We demonstrate the efficacy of our measure with applications to synthetic networks and to three real world hypergraphs, and compare our results with those obtained by applying other scalar measures of centrality proposed in the literature.

© 2022 Elsevier Ltd.

### 1. Introduction

Ranking nodes in a graph is certainly the most fundamental task in modern network science [1–4]. Already in 1977, Linton C. Freeman gave the first definition of betweenness centrality, and used it to rank individual clout in social networks [5,6]. The earliest definition and use of eigenvector centrality can even be traced more than a century ago, in 1895, when Edmund Landau used it for scoring chess tournaments [7]. Nonetheless, it was not before the discovery of heterogeneity in the degree distributions of real world networks [8] that the full depth

of implications of node centrality was realized. The 'hub' became, and still is, a popular meme that stands for influence, importance, or virality in social, biological and technological networks [1,9–12]. The identification of the most central nodes in complex networks is crucial for error and attack tolerance [13,14], viral marketing [15], information spreading [16–18], influence maximization [19,20], as well as plant genomic [21] and cancer research [22,23], just to name but a few examples. Not to mention that companies like Google are actually building their entire business in providing efficient and customized rankings of webpages.

Although the relevance of quantifying node centrality is undisputed, the best measure for it very much depends on the particularities of the problem at hand. The various measures adopted so far to quantify node centrality, from the simplest node degree to the variations of betweenness and eigenvector centrality [24–27], do not optimize a global

\* Corresponding author.

E-mail address: [karin.alfaro@urjc.es](mailto:karin.alfaro@urjc.es) (K. Alfaro-Bittner).

<sup>1</sup> These three authors equally contributed to the Manuscript.

function of influence, and are thus inherently unable to guarantee optimal performance [19]. Therefore, the correct question one has to ask himself is not how central is a given node in a network, but rather how central is a given node in a network *with respect to a given process*.

The issue is further exacerbated by the recent departures from traditional networks towards multilayer and higher-order networks as more apt representations of real world systems [28–30]. Although a generalization of eigenvector centrality for multiplex networks has been proposed [31], this does not account for the fact that in higher-order networks a link can connect more than two nodes. The potential of higher-order interactions has been recognized already in the early 70s by Ronald H. Atkin [32], but the interest peaked only recently with the inability of classic graph representations to describe group interactions. This ineptitude comes to a head when studying peer pressure, public co-operation, complex contagion or opinion formation, to list just a few examples that clearly extend well beyond dyadic interactions in social science [33,34], or when considering three or more species that routinely compete for food and territory in a complex ecosystem [35], or when functional [36] or structural [37] brain networks or protein interaction networks [38] are studied. Several different approaches to define higher-order interactions have been considered in the literature (see the review works [30,39–41] for a comprehensive account of the different definitions adopted for higher-order interactions). In specific instances, higher-order interactions have been modeled as (directed) sequences of nodes, sequences of set interactions, or motifs of dyadic edges, or combinations of two or several of those types [39–41]. In our work, we adopt what is possibly the most common notion of higher-order networks, i.e., is that of hypergraphs or simplicial complexes, where interactions between nodes are represented by a generalization of edges to hyperedges which capture (undirected and unweighted) group interactions. The generalization of our study to other settings will be considered in future works, in the line of what was recently suggested in Refs. [42–44].

In view of these recent developments, it is therefore crucial to generalize centrality measures in a way that they can account for higher-order interactions. In fact, some measures have been introduced in the literature that extends the classic notion of centrality to hypernetworks [45–48], but they all compute a single (scalar) number per node. In our study, we consider the most general case of an ensemble of  $N$  nodes which interplay by means of interactions of any order  $d \leq D$  (with  $D$  indicating the maximum order of group interactions taking place in the ensemble), and introduce instead a measure of centrality which is a *vector* assigned to each node, with dimension  $D - 1$ . While our vector centrality is related to the classical eigenvector centrality for networks containing only dyadic interactions, we will demonstrate that our measure has instead a significant added value (if compared with scalar measures) in all situations where interactions occur at higher-orders. We will show with practical applications that our measure is, indeed, able to distinguish different roles which may be played by a same node at different orders of interactions, a feature which is evidently impossible to be revealed by any single scalar measure.

### 1.1. The vector centrality measure

Let us start with considering  $N$  nodes which are interplaying by means of  $l_2$  links (dyadic interactions),  $l_3$  hyperlinks of order 3 (triadic interactions),  $l_4$  hyperlinks of order 4 (quadratic interactions), and in general by  $l_d$  hyperlinks of order  $d$  (with  $d = 2, 3, \dots, D$ ). We here concentrate on the case where all such hyperlinks are undirected. Mathematically this defines an undirected higher-order network (or hypergraph)  $G = (V, \mathcal{E})$ , i.e., a finite set  $V$  containing  $N$  nodes, and a family  $\mathcal{E}$  of  $\ell = \sum_{i=2}^D l_i$  non-empty and non-singleton subsets of nodes of  $G$ , each subset defining a hyperlink.

Our idea is to associate to  $G$  its linegraph  $L(G)$ , as introduced by Hassler Whitney for graphs in 1932 [49] and extended for higher-order

networks by Jean-Claude Bermond et al. in 1977 [50,51]. In particular,  $L(G)$  is a graph of  $l$  nodes (each of which mapping one of the hyperedges of  $G$ ). The links of  $L(G)$  stand for adjacency between hyperedges in  $G$ : if  $h_1 \in \mathcal{E}$  and  $h_2 \in \mathcal{E}$  are two hyperlinks, then there is an undirected link in  $L(G)$  between the nodes  $h_1$  and  $h_2$  if and only if  $h_1 \cap h_2 \neq \emptyset$ .

Fig. 1 depicts an illustrative example, where a hypergraph  $G = (V, \mathcal{E})$  is defined by  $V = \{1, 2, 3, 4, 5\}$  and  $\mathcal{E} = \{\{1, 2, 3\}, \{1, 2, 3, 4\}, \{2, 4, 5\}, \{4, 5\}\}$ . The figure shows also the associated linegraph  $L(G)$ , and the projection network  $\pi_2(G)$  of  $G$ . Notice that the projection of an hypergraph into a graph can be constructed in different ways. For instance, it can be defined as an unweighted network (see Ref. [30]) given by the number of hyperedges between two nodes [52], or as a weighted network [53], among others formalisms. In our work, the projection network  $\pi_2(G)$  of the hypergraph  $G = (V, \mathcal{E})$  is defined as the classic undirected and unweighted network formed by the same set of nodes as in  $G$  and whose links represent the dyadic interactions resulting from the projection of the hyperlinks of  $G$ .

Now, it is straightforward to demonstrate that if  $G$  is undirected and connected, then also  $L(G)$  is undirected and connected. Indeed, for any pair of hyperedges  $h_i$  and  $h_j$  in  $L(G)$  a path can be constructed by choosing a node  $v$  from  $h_i$  and a node  $w$  from  $h_j$ , and by using the same sequence of hyperedges as in path from  $v$  to  $w$  in  $G$ . Then, the classic Perron-Frobenius theorem [54,55] guarantees the existence and uniqueness of the eigenvector centrality of  $L(G)$ . In other words, one can compute with standard methods the classical eigenvector centrality of each node in  $L(G)$ , and one obtains a value  $c(h) \in [0, 1]$  for all hyperlinks  $h \in \mathcal{E}$  in  $G$ , such that  $\sum_{h \in \mathcal{E}} c(h) = 1$ .

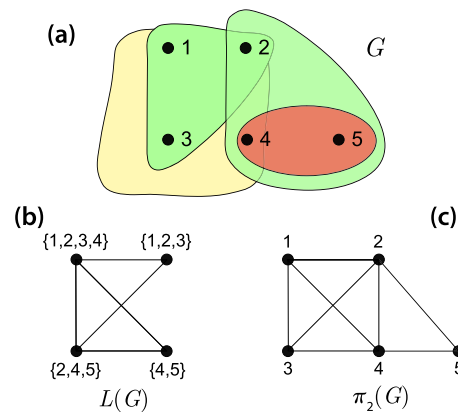
With the  $l$  values of  $c(h)$  at hand, we can now define the *vector centrality* of each node  $i \in V$ , a non-negative vector  $\vec{c}_i = (c_{i2}, \dots, c_{iD}) \in \mathbb{R}^{D-1}$  such that, for every  $2 \leq k \leq D$  one has

$$c_{ik} = \frac{1}{k} \sum_{\substack{h \in \mathcal{E} \\ |h| = k}} c(h), \quad (1)$$

where  $|h|$  indicates the order (or size) of the hyperedge  $h$ , and  $D = \max\{|h|; h \in \mathcal{E}\}$  is the maximal size of hyperedges in  $G$  (the maximal order of the group interactions affecting the  $N$  nodes in the ensemble).

In other words, the  $k^{\text{th}}$  component  $c_{ik}$  of the vector centrality of node  $i$  is the sum of the centralities of all hyperlinks of size  $k$  that contain  $i$  as one of the incident nodes, and the weight value  $\frac{1}{k}$  makes that

$$\sum_{i \in V} \|\vec{c}_i\|_1 = \sum_{i \in V} \sum_{k=2}^D c_{ik} = \sum_{i \in V} \sum_{k=2}^D \sum_{\substack{h \in \mathcal{E} \\ |h| = k}} \frac{c(h)}{k}. \quad (2)$$



**Fig. 1.** An illustrative example of a linegraph  $L(G)$  [panel (b)] of a higher-order network  $G = (V, \mathcal{E})$  with five nodes [panel (a)] and its projection network  $\pi_2(G)$  [panel (c)]. See text for specifications.

Now, for each  $i \in V$ , one has that

$$\sum_{k=2}^D \sum_{i \in h \in \mathcal{E}} \frac{c(h)}{k} = \sum_{i \in h \in \mathcal{E}} \frac{c(h)}{|h|}$$

because in the last double summation each hyperlink  $h \in \mathcal{E}$  such that  $i \in h$  appears exactly once. Therefore, by using this last expression in Eq. (2), and by summing over all nodes  $i \in V$ , one gets that

$$\sum_{i \in V} \|\vec{c}_i\|_1 = \sum_{i \in V} \sum_{h \in \mathcal{E}} \frac{c(h)}{|h|} = \sum_{h \in \mathcal{E}} \sum_{i \in h} \frac{c(h)}{|h|} = \sum_{h \in \mathcal{E}} c(h) = 1.$$

This latter expression has been obtained by simply changing the summation order, and taking into account that every summand  $\frac{c(h)}{|h|}$  appears exactly  $|h|$  times. The final result is, therefore, that  $\sum_{i \in V} \|\vec{c}_i\|_1 = 1$ , which implies that our measure is properly normalized.

Notice that if  $D = 2$ , i.e., only dyadic interactions exist in  $G$ , then for each node  $i$  one has  $\vec{c}_i = (c_{i2}) \in \mathbb{R}$ , where the scalar value  $c_{i2}$  is related with the  $i^{\text{th}}$  component ( $c_i'$ ) of the classic eigenvector centrality of  $G$ , as it was proved in [56]. Precisely, for  $D = 2$ , calling  $\lambda_1$  and  $\lambda_2$  the greatest and second greatest eigenvalue of the adjacency matrix of  $G$ , and denoting by  $\Delta$  the norm of the difference between our measure and the eigenvector centrality, Ref. [57] gave the following bounding relationship:

$$\Delta \equiv \sqrt{\sum_{i=1}^N (c_{i2} - c_i')^2} \leq \frac{(4 - \sqrt{2})^{\frac{1}{4}} \sqrt{N}^{\frac{1}{4}} \sqrt{2l_2}}{\lambda_1 - \lambda_2} \sqrt{\lambda_1 - \frac{2l_2}{N}},$$

which holds as far as the so called graph irregularity  $I(G) = \lambda_1 - \frac{2l_2}{N}$  is smaller than  $\left(\left(\sqrt{2} - 1\right)^2 / \left(4\sqrt{2} - 2\right)^4 N^2 \sqrt{2l_2}\right) (\lambda_1 - \lambda_2)^4$  (see Ref. [57] for details). We remark that this expression shows analytically that, for  $D = 2$ , i.e., when only dyadic interactions exist in  $G$ ,  $c_{i2}$  is very close to the  $i^{\text{th}}$  component ( $c_i'$ ) of the classic eigenvector centrality of  $G$ , and actually the more regular the network is the closer  $c_{i2}$  is to  $c_i'$ . In fact, it can actually be demonstrated that a more complicated construction of  $L(G)$  (where the setting of both nodes and links would account also for all possible permutations in the order of the nodes forming hyperlinks in  $G$ ), would lead to recover exactly (at  $D = 2$ ) the classical eigenvector centrality. For all practical purposes of this article, however, such a “directed” construction leads to the same qualitative results and quantitative rankings, and therefore we decided to report it elsewhere as a mathematical extension of our measure.

Moreover, we would like to remark here that the same entire procedure can be actually used for extensions of other structural measure (like, for instance, node betweenness) to higher-order networks: given a hypergraph  $G$ , one can always construct the associated linegraph, calculate the measure values for all hyperedges, and then use expressions similar to (1) to define vectorial quantities associated to the nodes in  $G$ .

## 2. Results

We here consider several practical examples to illustrate the added value of our vectorial centrality in distinguishing different roles a given node may have with respect to processes which may occur on top of interactions of different orders, a capacity which is instead greatly dwindled, if not prevented at all, using classical measures on the (weighted or unweighted) projections of the hypergraph.

### 2.1. Synthetic networks

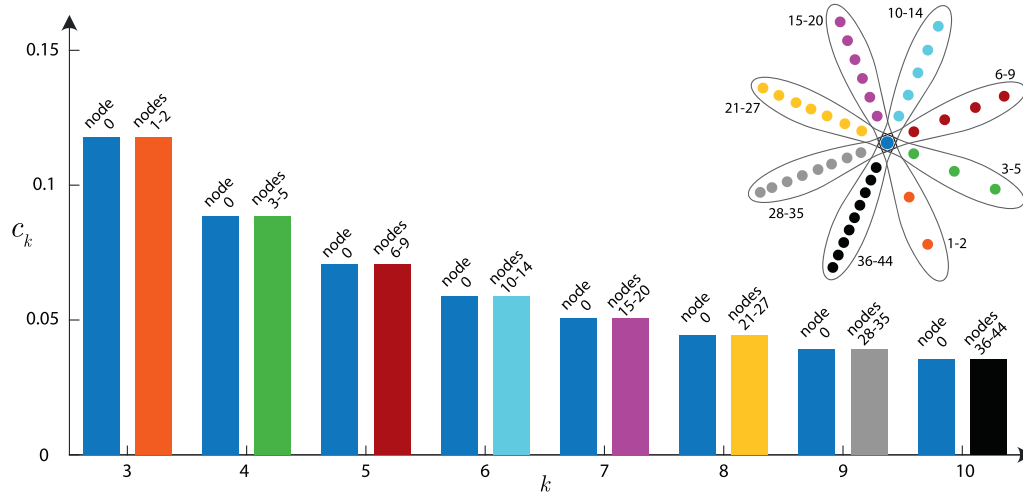
In order to provide a first comparison between our vectorial measure and the scalar centralities proposed so far, let us refer to the measures introduced in Refs. [45–48].

Reference [47] extends the methodology of eigenvector centrality for the case of simplicial complexes. A direct comparison with our measure is therefore not possible, as in simplicial complex one has to assume that the existence of a  $d$ -simplex (a simplex of order  $d$ ) automatically implies the existence of all possible interaction orders from 2 to  $d - 1$ , which is not the case for many real world higher-order networks, and which makes that framework totally different from the more general case of hyper-networks considered here.

In Ref. [46] three possible generalizations of eigenvector centrality for regular hypergraph are presented. The evident difference between our measure and the ones proposed in [46] is therefore the fact that we do not restrict hypergraph to be regular. However, even if we limit ourselves to the case of regular hypergraph, our results differ from the ones presented in [46]. In particular, let us refer to the same example that was made in Ref. [46], the so called sunflower hypergraph, a star-like hypergraph having one central node and  $r$  edges (sunflower's petals), each one of order  $d$ . In this example, our measure directly provides the ratio between the value of centrality of the central node,  $c_{0d}$ , and that of every other node  $i$ ,  $c_{id}$ . One indeed has that  $c_{0d} = \frac{1}{d} \sum_{0 \in h \in \mathcal{E}, |h|=d} c(h) = \frac{rc}{d}$  and  $c_{id} = \frac{1}{d} \sum_{i \in h \in \mathcal{E}, |h|=d} c(h) = \frac{c}{d}$ ,  $i \neq 0$ , with  $c$  being the line graph nodes' centrality (which, in this case, is equal for each node in  $L(G)$ , as the line graph is a clique of  $r$  nodes). It immediately follows that  $c_{0d}/c_{id} = r$ , which makes a strong difference with respect to what reported in Fig. 1 of Ref. [46].

Finally, Ref. [48] suggests a generalization of the famous HITS algorithm to hypergraphs, in which the nodes are more central if they are connected with more central hyperedges, and vice versa. This approach shares, indeed, similarities with the idea proposed in our study, as the components of our vector centrality are calculated from the values of the hyperedges' centralities in the line graph. However, the information that one can extract from the two measurements is completely different. For instance, let us analyze the same example presented in Ref. [48], i.e., a sunflower hypergraph with eight petals, which however correspond now to hyperedges of different sizes (from 3 to 10, see the pictorial sketch at the left of Fig. 2). In Ref. [48] node 0 has the highest centrality value, no matter which function (linear, max, log-exp) is used in the process of centrality calculation. For all other nodes, in the linear case (the log-exp case) the higher is the order of the hyperedge to which they belong the higher (the lower) is the value of the centrality, while in the max case the centrality values are all equal. In comparison, our vectorial measure provides a much richer information, as one can immediately see from Fig. 2. The central node's properties are now clearly distinguishable from those of all other nodes, primarily because it is the only one having non-zero values in all its centrality's components, whereas all other nodes feature a *localized* centrality value in the component corresponding to the order of the hyperedge to which they are belonging. Furthermore, at each hypergraph order, all the nodes in the corresponding petal have the same centrality component, once again differentiating our results from those of Ref. [48].

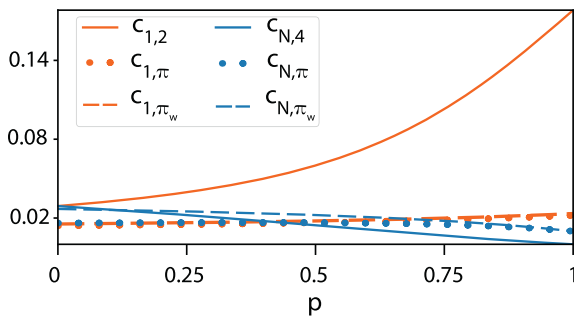
In order to better illustrate the qualities of our vectorial measure, let us now move to a more complicated synthetic hypergraph consisting of  $N = 100$  nodes, 400 hyperlinks of order 2, 400 hyperlinks of order 3 and 400 hyperlinks of order 4 (i.e.,  $l_2 = l_3 = l_4 = 400$ ), mapping therefore an ensemble of units interplaying by means of dyadic, triadic, and quadratic interactions. Here, we want to highlight how our vectorial centrality outperforms classical measures in tracking the importance of nodes when changes occur in the network structure. To this purpose, we initially prepare a graph with all  $l_3$  hyperlinks of order 3 which are randomly distributed. As instead for the  $l_2$  links of order 2 (the  $l_4$  hyperlinks of order 4), 350 of them are distributed randomly,



**Fig. 2.** The values of all components of the vector centrality for the nodes of a sunflower hypergraph with eight petals, each one corresponding to a hyperedge of different size (from 3 to 10, as it is seen in the pictorial sketch at the right of the figure, which has to be regarded also for the color code of the different bars appearing in the main plot). Our measure allows to clearly distinguish the properties of the central node's from those of all other nodes, as node 0 is the only one having non-zero values in all its components, whereas the centrality of all other nodes is *localized* only in the component corresponding to the order of the hyperedge to which they are belonging.

whereas 50 of them are placed so as to make vertex 1 (vertex 100) a hub for dyadic (quadratic) interactions, i.e., they are constructed so as to include vertex 1 (vertex 100) as one of the incident nodes. Then, we simulate limitations processes in group interactions by removing at random a fraction  $p$  of quadratic interactions, and we keep track on how the different centrality measures are efficient in monitoring the change of relevance of each node following the changes in the network structure. Precisely, one surveils the behavior of  $c_{1,2}$  (the first component of the vector centrality of node 1), of  $c_{N,4}$  (the last component of the vector centrality of node 100),  $c_{1,\pi}$  and  $c_{N,\pi}$  (the first and last components of the classical eigenvector centrality for the projected graph,  $\vec{c}_\pi$ ), and  $c_{1,\pi_w}$  and  $c_{N,\pi_w}$  (the first and last components of the classical eigenvector centrality for the projected weighted graph,  $\vec{c}_{\pi_w}$ ). In  $\pi$  two nodes are connected if there exists a hyper-link to which they both belong to; in  $\pi_w$  the weight of each link is the number of hyperlinks to which the two end nodes are belonging to.

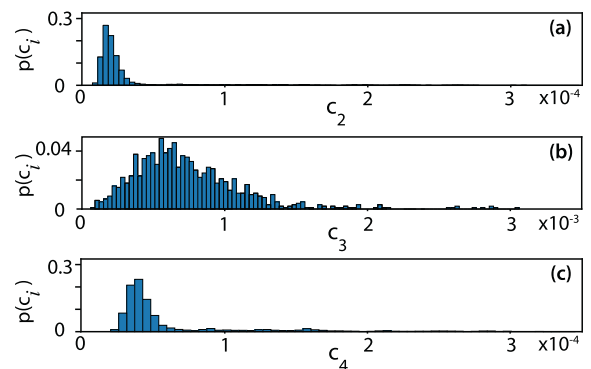
The results are reported in Fig. 3, and show clearly that only our vectorial measure (by comparison of  $c_{1,2}$  and  $c_{N,4}$ ) is able to reveal a substantial loss of centrality of node 100 as the number of quadratic interactions is progressively reduced, and a corresponding gain in centrality of node 2 which eventually remains the only hub in the system.



**Fig. 3.**  $c_{1,2}$ ,  $c_{N,4}$ ,  $c_{1,\pi}$ ,  $c_{N,\pi}$ ,  $c_{1,\pi_w}$  and  $c_{N,\pi_w}$  (see text for definition) vs. the fraction  $p$  of removed quadratic interactions, for the first synthetic network described in the text. The color code of the different curves is reported in the legend. Each point corresponds to an ensemble average over 5000 simulations: 100 different network realizations and for each one of them 50 different realizations of random removal of the  $l_4$  hyperlinks.

In a second example of a synthetic network, we probe the capability of our measure to reveal different scaling properties which may affect different orders of interactions in the graph, even in the case in which, at variance with the previous case, such orders do not correspond to the same number of hyperlinks. To that purpose, we construct another synthetic graph with dyadic, triadic, and quadratic interactions, this time with  $N = 1,000$ ,  $l_2 = 4,000$  hyperlinks of order 2,  $l_3 = 1,000$  hyperlinks of order 3, and  $l_4 = 2,000$  hyperlinks of order 4. All  $l_2$  and  $l_4$  hyperlinks are chosen randomly, this way determining a strongly homogeneous distribution for dyadic and quadratic interactions. Instead, the  $l_3$  hyperlinks (which are in the minimum number with respect to all other hyperlinks) are chosen so as to determine a strongly heterogeneous distribution: at each time those hyperlinks are constructed with a probability which explicitly depends on the actual node degree.

It has to be remarked that any projection (weighted or unweighted) of such synthetic higher-order network would result in a heterogeneous degree distribution, with the consequence that any centrality measure applied to such projected graph would reveal a strong heterogeneity. The results of applying our measure are, instead, shown in Fig. 4, where we report the histograms of the first ( $c_2$ , panel a), the second ( $c_3$ , panel b) and the third ( $c_4$ , panel c) component of our vectorial



**Fig. 4.** (a–c) Histograms (sampled with 100 bins) of the first ( $c_2$ , panel a), the second ( $c_3$ , panel b) and the third ( $c_4$ , panel c) component of the vectorial centrality, calculated over the second synthetic network described in the main text. Each histogram refers to values which are ensemble averaged over 10 different graph realizations.



centrality. It is seen that while the histograms reveal strong homogeneity at the level of dyadic and quadratic interactions (panels a and c), they clearly show heterogeneity traits at the level of triadic interactions (panel b), this way accounting exhaustively for the overall structural properties which have been engineered in the hypergraph.

## 2.2. Real-world hypergraphs

Finally, we calculate our measure on several real-world hypergraphs and discuss the added value of our measure in revealing important information on the structure of the considered hypernetworks.

The first considered hypergraph is that mapping the information publicly available in the arXiv (<https://arxiv.org/>, <https://github.com/mattbierbaum/arxiv-public-datasets/>) database, with the data parsing made by Ref. [58]. In particular, we focus on the data of preprints published in mathematics, and extract those papers which were written in collaboration, i.e., those having at least two co-authors. The extracted dataset consists of a total of 498,071 papers co-written by 230,605 authors.

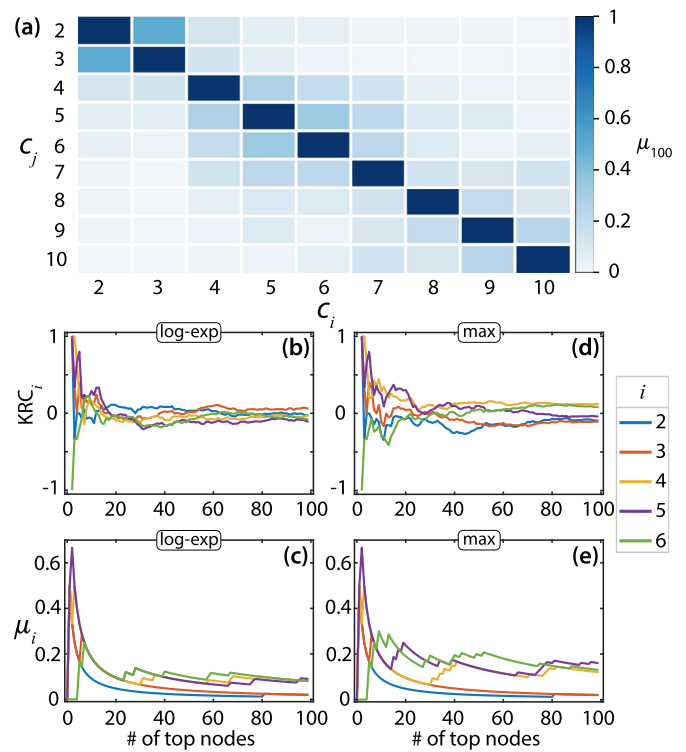
The data were mapped into a hypergraph  $G_{math}$ , where nodes were scientists, and each paper formed a hyperlink (a group interaction) of length equal to the number of co-authors. The maximal number of co-authors of a single papers (i.e., the maximal length  $D$  of hyperlinks in  $G_{math}$ ) is 67, which implies that the vectorial centrality of each scientist will have 66 components. The associated linegraph  $L(G_{math})$  is rather large in size: it is obviously formed by 498,071 nodes, and it has 9,808,188 links. The eigenvector centrality of  $L(G_{math})$  is then calculated, and the vector centrality of each scientist in  $G_{math}$  is evaluated.

Various rankings of scientists may be extracted according to the different components of the vector centrality, i.e., scientists may have different role and importance with respect to different hyperedges' sizes. In particular, we here analyze how many of the members of the top  $x$  authors's list in the ranking with respect to a given component of the vector centrality is also belonging to the top  $x$  authors's list in the ranking made with respect to another component. To do so, we introduce the fraction  $\mu_x$  as follows:

$$\mu_x(c^i, c^j) = \frac{|\text{top}_x(c^i) \cap \text{top}_x(c^j)|}{x}, \quad (3)$$

where  $c^i$  and  $c^j$  are, respectively, the  $i^{\text{th}}$  and  $j^{\text{th}}$  components of the vector centrality of the nodes,  $\text{top}_x(c^i)$  ( $\text{top}_x(c^j)$ ) is the set of the nodes which are occupying the top  $x$  positions in the ranking made by comparing the  $i^{\text{th}}$  (the  $j^{\text{th}}$ ) component of their vector centralities, and  $|\cdot|$  stands here for the cardinality of the set.  $\mu_x(c^i, c^j)$  measures therefore how large is the overlap between the two sets, and its values  $\mu_x(c^i, c^j)$  form a square matrix of  $66 \times 66$  elements, which actually describes how correlated are the positions scientists are holding in the ranking calculated with respect to a given component of the vector centrality with those held by the same scientists in the ranking calculated with respect to another component.

The values of  $\mu_{100}(c^i, c^j)$  (limited to the first ten components, out of the 66, of the vector centrality) are reported in panel a) of Fig. 5. It is seen that, except for the few values close to  $i = j$ , the fractions  $\mu_{100}(c^i, c^j)$  are relatively small for  $i \neq j$ , and therefore, the lists of the 100 top leaders in the rankings made with respect to different hyperedges sizes are significantly different. This confirms that the use of our measure is essential for extracting information on such differences, which would be instead inaccessible by any other scalar measure of centrality. In panels (b–e) of Fig. 5 a comparison is made with the unique ranking obtained by the use of the scalar measure of centrality proposed in Ref. [48] with a log-exp (panels b and d) and a max (panels c and e) function. This is done by reporting two different correlation measures: the Kendall rank correlation (KRC) coefficients (panels b and c) and the values of the function  $\mu$  as calculated by Eq. (3) when

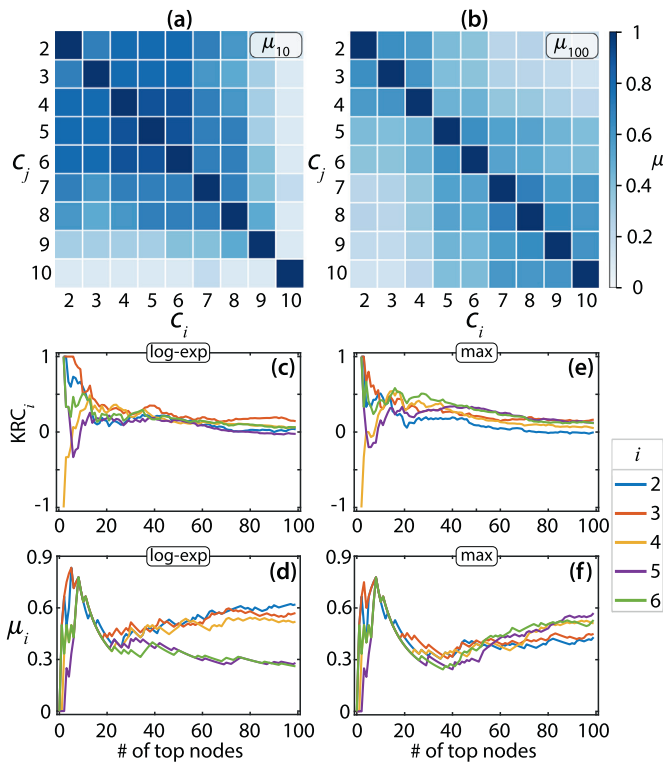


**Fig. 5.** (a)  $\mu_{100}(c^i, c^j)$  (see Eq. (3) of the text for definition) for the hypergraph representing scientific co-authorship in mathematics. Reported values are limited to the first ten components, out of the 66, of the vector centrality. It is clearly seen that, in general, the values of  $\mu_{100}$  are rather small for  $i \neq j$ . (b–e) Correlation between the rankings provided by the different components (from the second to the sixth, see color code in the legend at the right of the figure) of our vector centrality and the unique ranking obtained by adopting the algorithm of Ref. [48] with a log-exp (b,d) and a max (c,e) function. Panels b and c report the Kendall rank correlation (KRC) coefficients, while panels d and e report the values of the function  $\mu$  (from the same Eq. (3), in which  $c^i$  are substituted with the values of centralities extracted with the algorithm of Ref. [48]).

the values  $c^i$  are substituted with those extracted with the algorithm of Ref. [48] (panels d and e). Both correlation measures are reported as functions of the length of the ranking list. Lines of different colors correspond to different hyper-edge orders in our vector centrality measure. It is clearly seen that, when the ranking list is small in size, the intersections of the various sets of top ranked nodes is significant, implying that our measure individuates the same fundamental actor of the game. However, as the size of the ranking list increases, the corresponding Kendall rank correlation coefficients shrink, up to getting close to zero for every hyper-edge order. This implies that the obtained rankings do not differ substantially in individuating the really top nodes in the hypergraph, while they are fundamentally different as far as nodes of medium importance are considered.

A second real-world hypergraph is constructed from the data available at <https://www.cs.cornell.edu/arb/data/NDC-substances/> [59]. The dataset contains information on the composition of commercial drugs, posted by the U.S. Food and Drug Administration. In the dataset, each node represents a substance (or active principle, e.g., octinoxate, titanium dioxide, etc.) and each hyperlink stands for a commercial drug made of a given composition of such active principles. For the purposes of our application, drugs composed by no more than 25 substances were taken (the same as in Ref. [59]). Moreover, drugs consisting of only one active principle were excluded from the analysis. The result is an hypergraph consisting of 3438 nodes and 29,296 hyperlinks.

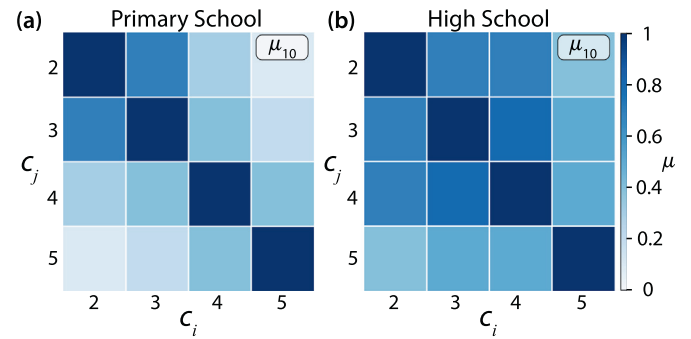
Panels a) and b) of Fig. 6 reports the results for  $\mu_{10}$  and  $\mu_{100}$ , as defined by Eq. (3). The information that our measure provides allows to infer that there are principles which are important for both the drugs with small number of ingredients and the ones with complex



**Fig. 6.** (a)  $\mu_{10}(c^i, c^j)$  and (b)  $\mu_{100}(c^i, c^j)$  (see Eq. (3) for definition) for the commercial drug higher-order network (all specifications of the hypergraph are given in the text). (c-f) Correlation between the rankings provided by the different components (from the second to the sixth, see color code in the legend at the right of the figure) of our vector centrality and the unique ranking obtained by adopting the algorithm of Ref. [48] with a log-exp (c,e) and a max (d,f) function. Panels c and d report the KRC coefficients, while panels e and f report the values of the function  $\mu$  (from the same Eq. (3), in which  $c^i$  are substituted with the values of centralities extracted with the algorithm of Ref. [48]).

composition [yielding a non negligible overlap between the top 10 ranked components for all hyperlink sizes in between 2 and 8, as can be seen in panel a) of Fig. 6]. However, in general [see panel b) of Fig. 6] the sets of key ingredients of the drugs with simple and complex compound are significantly different. Once again, we remark that such information would have not been extracted from scalar centrality measures. In panels (c-f) of Fig. 6 we again compare the ranking obtained by the different components of our vector centrality measure with the unique ranking obtained by the use of the scalar measure of centrality proposed in Ref. [48] with a log-exp (panels c and e) and a max (panels d and f) function. One can actually draw the same conclusions as above: when the list is small in size, the top ranked nodes are common to both rankings, but when the size of the ranking list increases the KRC coefficients get close to zero for every hyper-edge order.

As a third application, we consider the hypergraphs that can be constructed from the data on primary school contacts [taken from <https://www.cs.cornell.edu/arb/data/contact-primary-school/> [59,60]] and those on high school contacts [taken from <https://www.cs.cornell.edu/arb/data/contact-high-school/> [59,61]]. In both cases, data refer to experiments where wearable sensors, registering social interactions by proximity at a resolution of 20 s, are beard by students (242 kids in the case of the primary school, and 327 adolescents in the case of high school). As the dataset contains a lot of repetitions of the same group of people (the duration of the interactions are in general far larger than the 20 s resolution time), only unique groups were analysed. Furthermore, only edges with size no less than 2 were considered. The resulting hypergraph for primary school (high school) contacts consists therefore of 242 (327) nodes and 12,704 (7818) hyperlinks



**Fig. 7.**  $\mu_{10}(c^i, c^j)$  (see Eq. (3) for definition) for the primary [panel (a)] and high [panel (b)] school contacts hypernetwork (all specifications of the hypergraph are described in the text). It is seen (with particular evidence in the case of high school contacts) that social interactions pilot the emergence of *leaderships* of students which tend to be central independently on the group size.

corresponding to the total number of unique groups, i.e., the total number of nodes forming the line graph, which in its turn consists of a total number of edges of 2,238,167 (593,188).

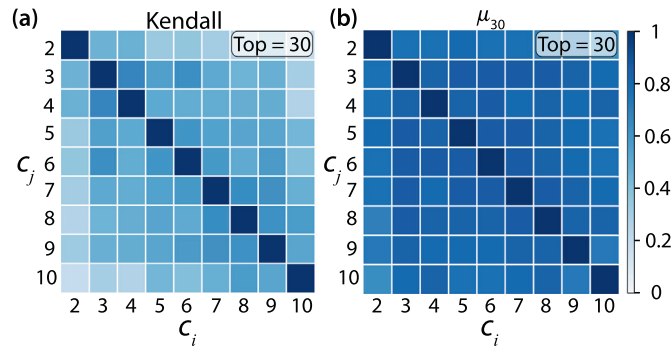
For both hypergraphs, the largest group size (the maximal order of interaction) is 5. Once again, to compare the ranking of students related to distinct edge sizes, we use the same measure  $\mu_k$ , defined by Eq. (3). The results are shown in Fig. 7). It is seen that *central* students in groups with size 2–4 are mainly not present in the top list of the groups of 5 people. However, it is seen (with particular evidence in the case of high school contacts) that such social interactions pilot the emergence of a *leadership* of students which tends to be central independently on the group size.

As the fourth application, we consider the hypergraph constructed from the data taken from <https://www.cs.cornell.edu/arb/data/senate-bills/> [62–64]. There, nodes are US Congress persons and hyper-links are co-sponsorships of bills which were put forth in the Senate. The dataset can be mapped into a hypergraph made of 294 nodes and 29,157 hyperedges, and the corresponding line graph has 29,157 nodes and 82,211,358 edges. After application of our method, Fig. 8 reports the KRC coefficients (panel a, calculated now for each pair of ranking lists, as obtained with the  $i^{th}$  and  $j^{th}$  components of the vector centrality) and  $\mu_{30}(c^i, c^j)$  (panel b, limited to the first ten components of the centrality vectors). From the plot  $\mu_{30}(c^i, c^j)$  it is rather evident that the leading roles are played always by the same actors (the parties' leaders), independently on the number of persons co-sponsoring the bill. Moreover, the orderings inside the leading groups are rather close to each other, as non-trivial KRC coefficients are obtained. Only the ranking with respect to hyperedges having size 2 seems to be weakly correlated with the others.

From the one hand this case is therefore rather different from all the previous ones, in that all components of the centrality vector are correlated rather strongly, and one could be tempted to say that there is no need here for the use of a vectorial measure. However, from the other hand it is only using our vector centrality that one can reveal that, in order to protect their leadership, central persons in political parties try to play key roles in groups of different sizes. In other words, a high correlation of the rankings related to different hyper-links orders gives also meaningful information.

Our final application is the hypergraph constructed with the data taken from <https://www.cs.cornell.edu/arb/data/walmart-trips/> [65]. In this hypergraph, nodes are products at Walmart and hyperlinks are sets of co-purchased products. It is composed by 88,860 nodes and 69,906 hyperedges, with a corresponding line graph made of 69,906 nodes and 33,046,972 edges.

The resulting KRC coefficients (panel a) and the values of  $\mu_{10}(c^i, c^j)$  (panel b) are presented in Fig. 9. It is seen that the intersections of the sets of top 10 products with respect to different orders are very high.



**Fig. 8.** KRC coefficients (panel (a), calculated for each pair of ranking lists obtained with the  $i^{\text{th}}$  and  $j^{\text{th}}$  components of the vector centrality) and  $\mu_{30}(c^i, c^j)$  (panel (b), limited to the first ten components of the centrality vectors) for the hypergraph reflecting bills' co-sponsorships in the US Senate (all specifications of the hypergraph are described in the text).

This means that there are some essential products which appear in each bill, no matter its sizes. However, when we analyze the set of top 100 products the sizes of these intersections progressively decrease, and the values of the KRC coefficients for vectors relating to the top 100 rankings are negligible (not shown in the figure). One can conclude that there exists a set of essential products which are bought frequently no matter which size the bill has. Other products are bought not so frequently, and their appearance in the bill is not highly determined by the bill size. Once again, we highlight that such kind of conclusions can be drawn only when the centrality measure has a vectorial character.

### 3. Discussion

Taken together, we have introduced a centrality measure able to overcome the inherent limitations of scalar centralities in higher-order networks. Our measure assigns a vector to each node, with dimension one lower than the dimension of the longest hyperlink in the network, and with every component thus determining the centrality of that node for a link with a particular length.

Our vector centrality is related to the classical eigenvector centrality for networks containing only dyadic interactions. Furthermore, by using artificially generated higher-order networks as well as data from real-world higher-order networks, we have demonstrated that our measure has instead a significant added value in all situations where interactions occur at higher-orders, in that it unveils different roles which may be played by a same node at different orders of interactions and therefore is the only one which accounts exhaustively for the properties of the overall interactive structure of the hypergraph. In particular, our measure gives a much richer information about centrality relationships than that extracted from other scalar measures recently introduced for hypergraphs.

As noted already when introducing our vector centrality, the same approach can be readily applied to other structural measures, which thus opens the path towards a wider applicability of our approach.

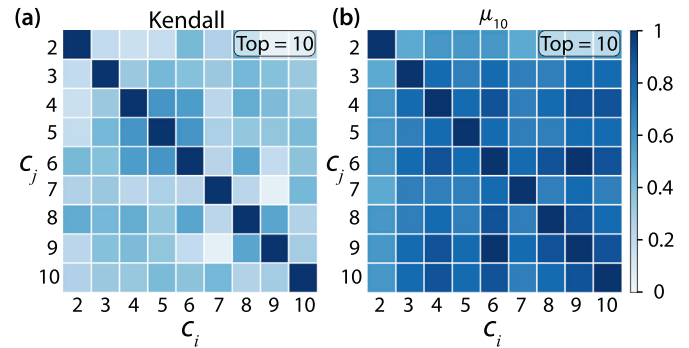
We expect our measure to become widely used with further progress in network science and related research fields.

### Data availability

Source data are available in the main text.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.



**Fig. 9.** KRC coefficients (panel (a), calculated for each pair of ranking lists obtained with the  $i^{\text{th}}$  and  $j^{\text{th}}$  components of the vector centrality) and  $\mu_{10}(c^i, c^j)$  (panel (b), limited to the first ten components of the centrality vectors) for the hypergraph constructed from the Walmart-tips dataset (all specifications of the hypergraph are described in the text).

### Acknowledgments

This work was supported by the Russian Federation Government (project "Post-crisis world order: challenges and technologies, competition and cooperation" funded by the Ministry of Science and Higher Education, agreement number 075-15-2020-783), by the program "Leading Scientific Schools" (Grant No. NSH-775.2022.1.1), by the Spanish Government (Project PGC2018-101625-B-I00 (AEI/FEDER, UE)), by the URJC Grant No. M1993, and by the Slovenian Research Agency (Grants No. P1-0403 and J1-2457).

### References

- [1] Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang D. Complex networks: structure and dynamics. *Phys Rep.* 2006;424:175–308.
- [2] Newman MEJ. *Networks*. Oxford, U.K.: Oxford University Press; 2010.
- [3] Jackson MO. *Social and economic networks*. Princeton, NJ: Princeton University Press; 2010.
- [4] Barabási A-L. *Network science*. Cambridge, U.K.: Cambridge University Press; 2016.
- [5] Freeman LC. A set of measures of centrality based on betweenness. *Sociometry*. 1977;40:35–41.
- [6] Freeman LC. Centrality in social networks conceptual clarification. *Soc Netw.* 1978;1:215–39.
- [7] Landau E. Zur relativen wertbemessung der turnierresultate. *Deutsches Wochensach.* 1895;11:3.
- [8] Barabási A-L, Albert R. Emergence of scaling in random networks. *Science*. 1999;286:509–12.
- [9] Albert R, Barabási A-L. Statistical mechanics of complex networks. *Rev Mod Phys*. 2002;74:47–97.
- [10] Newman MEJ. The structure and function of complex networks. *SIAM Review*. 2003;45:167–256.
- [11] Barthélemy M. Spatial networks. *Phys Rep.* 2011;499:1–101.
- [12] Holme P, Saramäki J. Temporal networks. *Phys Rep.* 2012;519:97–125.
- [13] Albert R, Jeong H, Barabási A-L. Error and attack tolerance of complex networks. *Nature*. 2000;406:378–82.
- [14] Cohen R, Erez K, Ben-Avraham D, Havlin S. Breakdown of the internet under intentional attack. *Phys Rev Lett.* 2001;86:3682.
- [15] Richardson M, Domingos P. Mining knowledge-sharing sites for viral marketing. In: *Proceedings of the eighth ACM SIGKDD international conference on knowledge discovery and data mining*; 2002. p. 61–70.
- [16] Kitsak M, Gallos LK, Havlin S, Liljeros F, Muchnik L, Stanley HE, Makse HA. Identification of influential spreaders in complex networks. *Nat Phys*. 2010;6:888–93.
- [17] Kleinberg JM. Authoritative sources in a hyperlinked environment. *46*; 1999. p. 604–32.
- [18] van den Heuvel MP, Sporns O. Network hubs in the human brain. *Trends Cogn Sci.* 2013;17:683–96.
- [19] Morone F, Makse HA. Influence maximization in complex networks through optimal percolation. *Nature*. 2015;524:65–8.
- [20] Lü L, Chen D, Ren X-L, Zhang Q-M, Zhang Y-C, Zhou T. Vital nodes identification in complex networks. *Phys Rep.* 2016;650:1–63.
- [21] Zanin M, Medina Alcazar J, Carbajosa JV, Gomez Paez M, Papo D, Sousa P, Menasalvas E, Boccaletti S. Parenclitic networks: uncovering new functions in biological data. *Sci Rep.* 2014;4:5112.
- [22] Karsakov A, Bartlett T, Ryblov A, Meyerov I, Ivanchenko M, Zaikin A. Parenclitic network analysis of methylation data for cancer identification. *PLoS ONE*. 2017;12:1–11.
- [23] Whitwell HJ, Blyuss O, Menon U, Timms JF, Zaikin A. Parenclitic networks for predicting ovarian cancer. *Oncotarget*. 2018;9:22717.

- [24] Brandes U. A faster algorithm for betweenness centrality. *J Math Soc.* 2001;25: 163–77.
- [25] Newman MEJ. A measure of betweenness centrality based on random walks. *Soc Netw.* 2005;27:39–54.
- [26] Piraveenan M, Prokopenko M, Hossain L. Percolation centrality: quantifying graph-theoretic impact of nodes during percolation in networks. *PLoS ONE.* 2013;8: e53095.
- [27] Alvarez-Socorro A, Herrera-Almaraz G, González-Díaz L. Eigencentrality based on dissimilarity measures reveals central nodes in complex networks. *Sci Rep.* 2015; 5:17095.
- [28] Boccaletti S, Bianconi G, Criado R, del Genio C, GómezGardeñes J, Romance M, Sendiña-Nadal I, Wang Z, Zanin M. The structure and dynamics of multilayer networks. *Phys Rep.* 2014;544:1–122.
- [29] Kivela M, Arenas A, Barthelemy M, Gleeson JP, Moreno Y, Porter MA. Multilayer networks. *J Complex Netw.* 2014;2:203–71.
- [30] Battiston F, Cencetti G, Iacopini I, Latora V, Lucas M, Patania A, Young J-G, Petri G. Networks beyond pairwise interactions: structure and dynamics. *Phys Rep.* 2020; 874:1–92.
- [31] Solá L, Romance M, Criado R, Flores J, García del Amo A, Boccaletti S. Eigenvector centrality of nodes in multiplex networks. *Chaos.* 2013;23:033131.
- [32] Atkin RH. From cohomology in physics to q-connectivity in social science. *Int J Man-Mach Stud.* 1972;4:139–67.
- [33] Burgio G, Matamalas JT, Gómez S, Arenas A. Evolution of cooperation in the presence of higher-order interactions: from networks to hypergraphs. *Entropy.* 2020;22:744.
- [34] Alvarez-Rodríguez U, Battiston F, de Arruda GF, Moreno Y, Perc M, Latora V. Evolutionary dynamics of higher-order interactions in social networks. *Nat Hum Behav.* 2021;5:586–95.
- [35] Levine JM, Bascompte J, Adler PB, Allesina S. Beyond pairwise mechanisms of species coexistence in complex communities. *Nature.* 2017;546:56–64.
- [36] Lord L-D, Expert P, Fernandes HM, Petri G, Van Hartevelt TJ, Vaccarino F, Deco G, Turkheimer F, Kringelbach ML. Insights into brain architectures from the homological scaffolds of functional connectivity networks. *Front Syst Neurosci.* 2016;10:85.
- [37] Sizemore AE, Giusti C, Kahn A, Vettel JM, Betzel RF, Bassett DS. Cliques and cavities in the human connectome. *J Comput Neurosci.* 2018;44:115–45.
- [38] Estrada E, Ross GJ. Centralities in simplicial complexes. Applications to protein interaction networks. *J Theor Biol.* 2018;438:46–60.
- [39] Torres L, Blevins AS, Bassett D, Eliassi-Rad T. The why, how, and when of representations for complex systems. *SIAM Review.* 2021;63:435–85.
- [40] Lambiotte R, Rosvall M, Scholtes I. From networks to optimal higher-order models of complex systems. *Nat Phys.* 2019;15:313–20.
- [41] Benson AR, Gleich DF, Higham DJ. Higher-order network analysis takes off, fueled by classical ideas and new data. 8 arXiv; 2021. 2103.05031.
- [42] Rosvall M, Esquivel AV, Lancichinetti A, West JD, Lambiotte R. Memory in network flows and its effects on spreading dynamics and community detection. *Nat Commun.* 2014;5:4630.
- [43] Scholtes I. When is a network a network? multi-order graphical model selection in pathways and temporal networks. *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '17.* New York, NY, USA: Association for Computing Machinery; 2017. p. 1037–46.
- [44] Zhao H, Xu X, Song Y, Lee DL, Chen Z, Gao H. Ranking users in social networks with motif-based pagerank. *IEEE Trans Knowl Data Eng.* 2021;33:2179–92.
- [45] Estrada E, Rodríguez-Velázquez JA. Subgraph centrality and clustering in complex hyper-networks. *Physica A.* 2006;364:581–94.
- [46] Benson AR. Three hypergraph eigenvector centralities. *SIAM J Math Data Sci.* 2019;1: 293–312.
- [47] Serrano DH, Gómez DS. Centrality measures in simplicial complexes: applications of topological data analysis to network science. *Appl Math Comput.* 2020;382:125331.
- [48] Tudisco F, Higham DJ. Node and edge nonlinear eigenvector centrality for hypergraphs, communications. *Physics.* 2021;4:1–10.
- [49] Whitney H. Congruent graphs and the connectivity of graphs. *Am J Math.* 1932;54: 150–68.
- [50] Bermond J-C, Heydemann M-C, Sotteau D. Line graphs of hypergraphs i. *Discrete Math.* 1977;18:235–41.
- [51] Heydemann M-C, Sotteau D. Line graphs of hypergraphs ii. *Colloq Math Soc J Bolyai.* 1976;18:567–82.
- [52] Carletti T, Battiston F, Cencetti G, Fanelli D. Random walks on hypergraphs. *Phys Rev E.* 2020;101:022308.
- [53] Banerjee A, Char A, Mondal B. Spectra of general hypergraphs. *Linear Algebra and its Applications.* 518. ; 2017. p. 14–30.
- [54] Perron O. Zur theorie der matrices. *Mathematische Annalen.* 1907;64:248–63.
- [55] Frobenius F. Über matrizen aus nicht negativen elementen. *Sitzungsberichte Akad Wiss Berlin.* 1912;26:456–77.
- [56] Criado R, Flores J, García del Amo A, Romance M. Analytical relationships between metric and centrality measures of a network and its dual. *J Comput Appl Math.* 2011;235:1775–80.
- [57] Criado R, Flores J, García del Amo A, Romance M. Centralities of a network and its line graph: an analytical comparison by means of their irregularity. *Int J Comput Math.* 2014;91:304–14.
- [58] Clement CB, Bierbaum M, O'Keefe KP, Alemi AA. On the use of ArXiv as a dataset. 19 arXiv; 2019. 1905.00075.
- [59] Benson AR, Abebe R, Schaub MT, Jadbabaie A, Kleinberg J. Simplicial closure and higher-order link prediction. *Proc Natl Acad Sci.* 2018;115:E11221–30.
- [60] Stehlé J, Voirin N, Barrat A, Cattuto C, Isella L, Pinton J-F, Quaggiotto M, den Broeck WV, Régis C, Lina B, Vanhems P. High-resolution measurements of face-to-face contact patterns in a primary school. *PLoS ONE.* 2011;6:e23176.
- [61] Mastrandrea R, Fournet J, Barrat A. Contact patterns in a high school: a comparison between data collected using wearable sensors, contact diaries and friendship surveys. *PLoS ONE.* 2015;10:e0136497.
- [62] Chodrow Philip S, Veldt Nate, Benson Austin R. Generative hypergraph clustering: From blockmodels to modularity. *Science Advances.* 2021;7(28):eabh1303.
- [63] Fowler JH. Connecting the congress: a study of cosponsorship networks. *Polit Anal.* 2006;14:456–87.
- [64] Fowler JH. Legislative cosponsorship networks in the US house and senate. *Soc Netw.* 2006;28:454–65.
- [65] Amburg I, Veldt N, Benson AR. Clustering in graphs and hypergraphs with categorical edge labels. *Proceedings of the Web Conference;* 2020.