Mathematics

Research article

Uplifting edges in higher-order networks: Spectral centralities for non-uniform hypergraphs

Gonzalo Contreras-Aso[1,2,*], **Cristian Pérez-Corral**[2,3] **and Miguel Romance**[1,2]

[1] Departamento de Matemática Aplicada, Ciencia e Ingeniería de los Materiales y Tecnología Electrónica, Universidad Rey Juan Carlos, 28933 Móstoles (Madrid), Spain

[2] Laboratorio de Computación Matemática en Redes Complejas y sus Aplicaciones, Universidad Rey Juan Carlos, 28933 Móstoles (Madrid), Spain

[3] Valencian Research Institute for Artificial Intelligence, Universitat Politècnica de València, Camino de Vera, s/n 46022 Valencia, Spain

* **Correspondence:** Email: gonzalo.contreras@urjc.es.

Abstract: Spectral analysis of networks states that many structural properties of graphs, such as the centrality of their nodes, are given in terms of their adjacency matrices. The natural extension of such spectral analysis to higher-order networks is strongly limited by the fact that a given hypergraph could have several different adjacency hypermatrices, and hence the results obtained so far are mainly restricted to the class of uniform hypergraphs, which leaves many real systems unattended. A new method for analyzing non-linear eigenvector-like centrality measures of non-uniform hypergraphs was presented in this paper that could be useful for studying properties of $\mathcal{H}$-eigenvectors and $\mathcal{Z}$-eigenvectors in the non-uniform case. In order to do so, a new operation—the *uplift*—was introduced, incorporating auxiliary nodes in the hypergraph to allow for a uniform-like analysis. We later argued why this was a mathematically sound operation, and we furthermore used it to classify a whole family of hypergraphs with unique Perron-like $\mathcal{Z}$-eigenvectors. We supplemented the theoretical analysis with several examples and numerical simulations on synthetic and real datasets: On the latter, we find a clear improvement over the existing methods, specially in cases where there is a huge disparity between the structure at each order, and on the former, we find that regardless of the chosen uniformization scheme, the nodes were similarly ranked.

1. Introduction

The last decade has seen a rise in the multidisciplinary field of "complexity science", partly due to the advances in computation, but also due to important milestones in theory. Within this far-reaching field, one of the most active and important areas is that of complex networks: The study of systems reduced to individuals and their interactions [1]. This area was originated in the mathematics of graph theory, but it was also supplemented by ideas and concepts from statistical physics, biology and computational science, to mention a few, yielding an ever-increasing amount of results in these and other areas, such as sociology, dynamical systems, and even multilinear algebra.

Theoretical advances in complex network theory have shed light on its own limitations; in particular, during these last years a lot of effort has been poured in extending the analytical, conceptual, and numerical tools already available for graphs to the realm of hypergraphs. Hypergraphs are natural extensions of graphs, where interactions are not restricted to be of a pairwise nature, i.e., one can have interactions between two, three, four, or even higher individuals. These are sometimes referred to as higher-order interactions, or hyperedges for short. The quintessential example of these kinds of systems is that of co-authorship networks: There, nodes represent authors and scientific articles are the interactions between them (which clearly are not restricted to papers written by pairs of authors).

Although there is already a plethora of results brought from classic complex network theory [2–4], there is an even higher amount of results which have yet to find their way into hypergraphs, to the point that it is a current area of research (see, for instance, [5]). The reason is pretty simple: Considering higher-order interactions usually complicates analytical calculations to the point where certain approximations or assumptions need to be established in order to make any progress at all. An illustrative example is that of obtaining spectral properties, such as spectral centrality measures, which provide a way to quantify the importance of nodes beyond their degrees, taking into account the relevance of their neighbors, and it is the basis of some real-world applications such as PageRank [6]. In graphs, the study of these properties translates to studying matrices (adjacency, Laplacian, etc.) whereas in hypergraphs, it translates to studying tensors, and while the former has been extensively studied, the latter has not.

The previous example relates to the subject of this paper. In order to generalize spectral centralities of graphs to the case of hypergraphs, Benson [7] made use of the recently developed theory of $\mathcal{H}$-eigenvectors and $\mathcal{Z}$-eigenvectors of tensors [8] in order to formulate a series of centrality measures. This formulation does apply, however, to uniform hypergraphs (those where the hyperedges are all of the same order), which strongly limits the application of the obtained results to many real systems that can only be modeled with hypergraphs with hyperedges of different sizes. The main goal of this paper is to present a method that allows considering non-linear eigenvector-like centrality measures of non-uniform hypergraphs and therefore extends the results obtained by Benson to uniform higher-order networks [7].

The technique we present is based on incorporating auxiliary nodes in a non-uniform hypergraph in order to enable a uniform-like analysis, such that we can use all the tools and results obtained for uniform hypergraphs, but now for a general one. This uniformization process, named the *uplift* of an hypergraph, is somehow the dual of the projection process that transforms hyperedges among k nodes into hyperedges among $2 \le j < k$ nodes and it has sound mathematical properties that make it very useful for analyzing several structural properties of hypergraphs and hypermatrices, as will be shown

along this paper.

It is important to point out that, although we will only focus on the analytical tools proposed and analyzed in this manuscript within the subject of network theory and centrality measures, the methods can be leveraged in all other areas where tensor eigenproblems have made an appearance, such as biology [9], medical imaging [10], quantum entanglement [11], data mining [12], or higher-order Markov chains [13].

This article is structured as follows. In Section 2, we provide the mathematical notions required throughout, as well as present a summary of the State of the Art. In Section 3, we define the *uplift* operation, which allows us to uniformize any hypergraph in a way that appropriately generalizes the uniform $\mathcal{H}$-eigenvector centrality, something which is then discussed in Section 4. We show why it fails to generalize the $\mathcal{Z}$-eigenvector centrality, although we rescue it in Section 5 to solve a problem in multilinear algebra, which in turn feeds back into the problem of $\mathcal{Z}$-eigenvector centrality of certain hypergraphs. Several examples and numerical computations on a variety of synthetic and real higher-order networks are included throughout the paper in order to illustrate the analytical results presented and compare them with other proposals and methods.

2. Preliminaries

We will start by giving a brief overview of the main concepts and notation which will be needed in what follows, first regarding the eigenvector centrality in standard graphs and then introducing useful hypergraph notions.

The eigenvector centrality in graphs: This work is framed in the context of centrality measures in networks [1]. A centrality measure is simply a way to provide a notion of importance to the nodes within a network, based on criteria (heuristics) regarding what constitutes importance. This criteria is subjective, and must be decided with an application in mind. The simplest criteria is counting the number of neighbors of each node. This is the so-called Degree Centrality. However, for many real-world applications (see [14] and the references therein), we are interested in also considering the importance of each neighbor in the computation of a node's importance. That is the basis for all spectral centralities.

The simplest spectral centrality in standard, pairwise interaction networks is the so-called Eigenvector Centrality. The heuristics behind it is the statement that in a graph $G = (V, E)$, a node's importance is proportional to the importance of its neighbors [15]. Mathematically,

$$c_i \propto \sum_{j \to i} c_j \Rightarrow \lambda \mathbf{c} = A^T \mathbf{c}, \tag{2.1}$$

where A is the adjacency matrix of the graph G. This is the well-known eigenvector equation of A, with $(\lambda, \mathbf{c})$ being the eigenvalue and eigenvector pairs. By the Perron-Frobenius Theorem, if the graph G is connected, then A is irreducible, and therefore the existence and uniqueness of a positive eigenvector $\mathbf{c}$ associated with the spectral radius $\rho(A)$ is guaranteed [16]. This theorem has provided support for a plethora of spectral centralities in standard complex network theory, out of which the Eigenvector Centrality is the paradigmatic example.

Notions from algebraic hypergraph theory: A hypergraph $H = (V, E)$ consists of a set of nodes $V = \{i_1, ..., i_n\}$ and a set of edges E. Each hyperedge consists of yet another set (or multiset, as will be discussed in Section 3) of nodes belonging to V. The size of a hyperedge is the number of elements within it. If the hypergraph is weighted, then there exists a function $w : E \to \mathbb{R}$, and $w(e)$ is the weight of edge e.

We say that a hypergraph is m-uniform if all of its hyperedges are of the same size m. Note that this subclass of hypergraphs is uncommon in real networks: If we consider the quintessential example of hypergraphs, which is the network of collaboration between scientists, the number of authors (nodes) in each hyperedge (paper) might not always be the same. Note also that the case of 2-uniform hypergraphs coincides precisely with networks of pairwise interactions.

Working with m-uniform tensors, if possible, is preferable as there are several analytical tools available for them. Namely, one can define the "hypergraph adjacency tensor*" $T \in \mathbb{R}^{[m,n]}$, whose components are

$$T_{i_1 \dots i_m} = \begin{cases} w(\{i_1 \dots i_m\}) & \text{if } \{i_1, \dots, i_m\} \in E, \\ 0 & \text{otherwise,} \end{cases} \quad 1 \le i_1, \dots, i_m \le n, \tag{2.2}$$

where m is the "order" of the tensor (equivalently, the size of its hyperedges), and $i_1, \dots, i_m \in V$. In the context of undirected hypergraphs, the tensors which we will be discussing will always be symmetric, meaning that $T_{i_1 \dots i_k} = T_{\sigma\{i_1, \dots, i_m\}}$, for all $\sigma \in \mathfrak{S}_m$, where $\mathfrak{S}_m$ denotes the permutation group of m indices.

The key ingredient connecting the Perron-Frobenius Theorem to the eigenvector centrality of graphs was the relation between the irreducibility of the adjacency matrix and the strong connectivity of the graph. In hypergraphs, generalizations of the Perron-Frobenius Theorem have been put forward for different tensor eigenproblems [8], which do require a similar connection to be made between irreducibility of the adjacency tensor and strong connectivity of the corresponding hypergraph.

Similarly to the matricial case, we can distinguish between reducible and irreducible tensors, but with a bit of a twist.

Definition 2.1 (Irreducible tensor [8]). *An order-m, dimension-n tensor T is reducible if there is a nonempty proper index subset $J \subset [n]$ such that*

$$T_{i_1 \dots i_m} = 0, \; \forall i_1 \in J, \; \forall i_2, ..., i_m \notin J. \tag{2.3}$$

If T is not reducible, then it is irreducible.

Here we used the notation $[n] = \{1, ..., n\}$. A number of results have been proven relating connectedness properties of hypergraphs to the irreducibility of this tensor [17, 8]. However, unlike the pairwise case, the intuitive notion of connectedness in a hypergraph does not directly translate to irreducibility of the associated tensor (see Example 2.7 of [18]). As it happens, tensor irreducibility is too strong a constraint to fully describe general hypergraphs.

Instead, strongly connected hypergraphs are described in terms of *weakly* irreducible tensors.

---

*The name "tensor" is usually reserved for mathematical objects invariant under coordinate transformations. In our case, we are instead referring to multidimensional arrays (or hypermatrices), which we refer to as tensors for the sake of conciseness.

Definition 2.2 (Weakly irreducible tensor [8])**.** *Let $M = (m_{ij})$ be a $n \times n$ non-negative matrix defined by*

$$m_{ij} = \sum_{j_3,\dots,j_m=1}^{N} |T_{ijj_3\dots j_m}|. \tag{2.4}$$

Then T is weakly irreducible if and only if M is irreducible.

This definition is equivalent to the intuitive notion of connectedness, as the graph whose adjacency matrix is M will have an edge between nodes i and j if there is at least a hyperedge containing them.

A hypergraph is said to be strongly connected if its adjacency tensor is irreducible in the previous sense. Luckily, most of the existence and uniqueness results which will be relevant for us have also been proven for these types of tensors [8].

Before moving on to the next subsection, let us define a ubiquitous operation involving a tensor $T \in \mathbb{R}^{[m,n]}$ and a vector $\mathbf{c} \in \mathbb{R}^n$, whose contraction produces yet another vector, sometimes called *tensor apply* [19] or TTSV1 [20]:

$$\boldsymbol{x} = T\mathbf{c}^{m-1} \iff x_{i_1} = \sum_{i_2\dots,i_m=1}^{n} T_{i_1 i_2\dots i_m}\mathbf{c}_{i_2}\dots\mathbf{c}_{i_m}. \tag{2.5}$$

2.1. Hypergraph spectral centralities: State of the Art

For simplicity, we will now restrict ourselves to 3-uniform hypergraphs, although the generalization to k-uniform hypergraphs is straightforward. The most straightforward generalization of the pairwise Eigenvector Centrality consists of defining functions $f : V \to \mathbb{R}$ and $g : V \times V \to \mathbb{R}$, and imposing the equation

$$f(c_i) = \frac{1}{\lambda} \sum_{\{i,j,k\}\in E} g(c_j, c_k). \tag{2.6}$$

The question then would be determining whether $\mathbf{c} = (c_1, \dots, c_n)^T$ exists and is unique, and if so, how could one calculate it.

So far in the literature, three choices of f, g have been considered [7], due to their simplicity, sensibility, and the existence of Perron-Frobenius-like associated theorems.

- **Clique motif Eigenvector Centrality (CEC)**: In this case $f(c_i) = c_i$ and $g(c_j, c_k) = c_j + c_k$. This choice is the simplest one, for it leads to a standard eigenvector equation, unlike in the next two cases.

$$c_i = \frac{1}{\lambda} \sum_{\{i,j,k\}\in E} (c_j + c_k). \tag{2.7}$$

This is tantamount to considering the standard (pairwise) Eigenvector Centrality of the motif adjacency matrix W of the hypergraph [7].
The main drawback of this approach is that it hides the higher-order nature of the hypergraph, reducing the problem of computing its centrality scores to that of a standard graph with a modified adjacency matrix.

- **$\mathcal{Z}$-Eigenvector Centrality (ZEC)**: In this case $f(c_i) = c_i$ and $g(c_j, c_k) = c_j c_k$. This choice fully incorporates the higher-order nature of the hypergraph by means of a non-linear g.

$$c_i = \frac{1}{\lambda} \sum_{\{i,j,k\} \in E} c_j c_k = \frac{1}{\lambda} \sum_{j,k=1}^{n} T_{ijk}\, c_j\, c_k \Rightarrow \lambda \mathbf{c} = 2T\mathbf{c}^2. \tag{2.8}$$

Remark 2.3. *Note that, if the eigenpair $(\lambda, \mathbf{c})$ is a solution to this equation, then for any $\alpha \in \mathbb{R}$ the eigenpair $(\alpha\lambda, \alpha\mathbf{c})$, is also a solution. This is problematic, as it means there are infinite eigenvalues. To deal with this, the unitarity condition $\mathbf{c}^T \mathbf{c} = 1$ is also imposed[†], although it makes the $\mathcal{Z}$-eigenvector $\mathbf{c}$ not rescalable.*

This equation was proven to always have a positive $\mathbf{c} > 0$ solution, provided the underlying hypergraph is strongly connected [22]. Moreover, this definition of eigenvectors, unlike the next one, is invariant under orthogonal transformations [8], making it physically relevant.

Remark 2.4. *This spectral problem features a different behavior depending on whether the order of the tensor T is even or odd. In the former case, for every eigenpair $(\lambda, \mathbf{c})$, the eigenpair $(\lambda, -\mathbf{c})$ is also a solution. In the latter case, for every eigenpair $(\lambda, \mathbf{c})$, the eigenpair $(-\lambda, -\mathbf{c})$ is also a solution. Therefore, in that case the spectral radius does not correspond to a unique eigenvalue, however for centrality purposes we are only interested in the ranking, and we can thus choose to keep just the positive solution.*

However, this measure has two important flaws: namely, the solution is not unique (even after imposing unitarity, see Example 2.7 of [18]), and all computational methods known to converge to a solution are computationally expensive [8].

- **$\mathcal{H}$-Eigenvector Centrality (HEC)**: In this case $f(c_i) = c_i^2$ and $g(c_j, c_k) = c_j c_k$. In this case we consider the same non-linear function g, but consider a function f which "dimensionally matches" the right-hand side (supposing centrality is measured in some "unit").

$$c_i^2 = \frac{1}{\lambda} \sum_{\{i,j,k\} \in E} c_j c_k = \frac{1}{\lambda} \sum_{j,k=1}^{n} T_{ijk}\, c_j\, c_k \Rightarrow \lambda \mathbf{c}^{[2]} = 2T\mathbf{c}^2, \tag{2.9}$$

where $\mathbf{c}^{[2]}$ refers to the Hadamard (componentwise) square of the vector $\mathbf{c}$.

This equation not only is guaranteed to have a positive $\mathbf{c} > 0$ solution if the hypergraph is strongly connected, but said solution is also unique (up to scaling) [22].

We will not further discuss the properties, advantages, and disadvantages of each of these methods. The interested reader is referred to [7] for an in-depth analysis of them in the context of hypergraphs, and to [8] for a general discussion of their mathematical properties, special cases, and results.

Up to this point we have limited the discussion to m-uniform hypergraphs. However, most real hypergraphs are not uniform, having edges of a variety of sizes. For instance, one of the prime examples of real hypergraphs is that of collaboration networks, where nodes represent researchers and hyperedges represent papers where their authors have collaborated. In this simple example it is clear that the corresponding hypergraph will contain pairwise interactions (papers with two authors), triple interactions (papers with three authors), and so on.

[†]Vectors satisfying $\mathbf{c}^T \mathbf{c} = \|\mathbf{c}\|_2 = 1$ (Euclidean norm) are sometimes referred to as $\mathcal{Z}_2$-eigenvectors in the literature [21, 8], as opposed to $\mathcal{Z}_1$-eigenvectors satisfying $|\mathbf{c}|_1 = 1$ (ℓ_1-norm).

2.1.1. Vectorial characterizations

A first idea on how to characterize the centrality of nodes in a non-uniform hypergraph is introducing the notion of a *vectorial centrality score*. For instance, in the HEC case, one could compute the centrality vector $\mathbf{c}^{(m)} \in \mathbb{R}^n$ associated with each order-m sub-hypergraph (provided they are all strongly connected), and then assign to each i node a vector $\mathbf{v}^i \in \mathbb{R}^{m-1}$ whose m-th component corresponds to its centrality score at order m, i.e., $v_m^i = c_i^{(m)}$.

The main drawback of the above approach is the fact that the hypergraph will likely not be strongly connected at each order (see, for instance, Figure 1). Not only that, but there interplay between different orders is completely absent. An attempt to palliate both problems was put forward in [23], where they resort to the line graph of a hypergraph (a structure which is proven to be strongly connected if the original hypergraph is as well) to translate the problem to that of hyperedge centrality scores, which can be tackled using standard, pairwise graph theory. These edge centralities are then "shared" among the nodes participating in each of them, at each level k, conforming a vectorial centrality score.

2.1.2. Embracing non-uniformity

Apart from the vectorial characterizations of the hypergraph centrality, it seems rather natural to wonder whether one can extend the notions of CEC, ZEC, and HEC to these non-uniform cases.

Following the traditional heuristics of the Eigenvector Centrality, the most general scenario would then be an equation of the form

$$\lambda f(c_{i_1}) = \alpha_2 \sum_{\{i_1,i_2\}\in E} g_2(c_{i_2}) + \alpha_3 \sum_{\{i_1,i_2,i_3\}\in E} g_3(c_{i_2}, c_{i_3}) + \cdots + \alpha_m \sum_{\{i_1,\ldots,i_m\}\in E} g_m(c_{i_2}, c_{i_3}, \ldots, c_{i_m}), \quad (2.10)$$

with $\alpha_2, \alpha_3, \ldots, \alpha_m > 0$ weighting factors assigned a priori, accounting for the relevance of each order of interaction.

If some of the functions $f, g_2, g_3, \ldots, g_m$ are non-linear, as in the HEC and ZEC cases, this equation is not known to have a solution in general. Not only that, but taking insight from the m-uniform case, it is clear that the choices of said functions will drastically change the existence and uniqueness properties of the solution. Our best shot at making progress in the non-uniform, non-linear case is then returning to the ZEC and HEC cases, and see if there is any way to introduce the non-uniformity there.

3. The uplift

We start with the uniformization of the hypergraph from the bottom up. For that, consider a hypergraph $H = (V, E)$ whose maximum hyperedge size is M, and a size $m \geq M$, possibly with multiset hyperedges, i.e., hyperedges where the same node is contained more than once (hypergraphs with such particularity are sometimes referred to as "multihypergraphs" [8]). We can turn every hyperedge of size lower than m into that size by adding an auxiliary node, which we name "$\star$", possibly multiple times within the same hyperedge.

Note that this notion of adding extra nodes is already present in other hypergraph-related works, although with completely different purposes: In [24] they call it "augmentation" and use it for community detection, in [25] they call it "inflation" and use it for hypergraph polynomials. In either case they use a simpler, unweighted version, which only encodes adjacency but not the strength of it, in contrast with our proposal.

More precisely, we have the following definition.

Definition 3.1 (Uplifted hypergraph at order m)**.** *Let $H = (V, E)$ be a hypergraph whose maximum hyperedge size is M and let $m \geq M$. We define the uplifted hypergraph at order m as*

$$\widetilde{H} = (\widetilde{V}, \widetilde{E}), \quad where \quad \widetilde{V} = V \cup \{\star\} \quad and \quad \widetilde{E} = \left\{ e \cup \left(\bigcup_{l=0}^{m-|e|} \{\star\} \right), \, e \in E \right\}. \tag{3.1}$$

Remark 3.2. *Note that, from a set-theoretic point of view, uplifted hyperedges are multiset objects, i.e., they may contain the auxiliary node $\star$ more than once.*

To exemplify this concept, Figure 1 illustrates the uplifting procedure with a simple case (a hypergraph with two 2-hyperedges and one 3-hyperedge uplifted to order 3 with an auxiliary node).
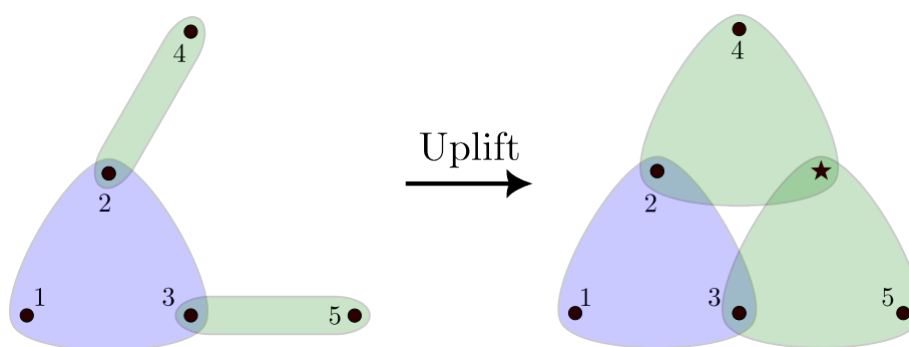

Figure 1. Uplift of the hypergraph $H = (\{1, 2, 3, 4, 5\}, \{\{1, 2, 3\}, \{2, 4\}, \{3, 5\}\})$ to $\widetilde{H}$.

The next step is constructing its associated tensor $\widetilde{T}$, in particular its components $\widetilde{T}_{i_1 i_2 \ldots i_m}$, which will be used in the centrality calculation. In order to do so, one can, naïvely, identify $\star$ as the node $n + 1$ and start filling in the entries of the tensor. There is a caveat, though: As we are considering undirected hypergraphs, the tensors at each order are considered symmetrized. Adding the extra node would provide more permutations to each hyperedge than those originally present. We can avoid this double counting by adding suitable combinatorial factors to the hyperedges which have been uplifted.

Taking this into account, we can define the uplifted tensor.

Definition 3.3 (Uplifted tensor of a hypergraph)**.** *Let $H = (V, E)$ be an unweighted, uniform hypergraph and let $\widetilde{H} = (\widetilde{V}, \widetilde{E})$ be its uplift to order m. The components of the uplifted tensor $\widetilde{T}_{i_1 i_2 \ldots i_m}$ associated with $\widetilde{H}$ are defined as*

$$\widetilde{T}_{i_1 i_2 \ldots i_m} = \begin{cases} 1 & if \quad \{i_1, i_2, \ldots, i_m\} \in E, \\ \dfrac{m^\star (m - m^\star)!}{m!} & if \quad \{i_1, i_2, \ldots, i_m\} \notin E \, \wedge \, \{i_1, i_2, \ldots, i_m\} \in \widetilde{E}, \\ 0 & otherwise, \end{cases} \tag{3.2}$$

where $i_s \in \widetilde{V}$, $s = 1, \ldots, m$, and with $m^\star$ being the number of times the auxiliary node $\star$ was added to the original hyperedge $e = \{j_1, \ldots, j_{m-m^*}\} \subset \{i_1, \ldots, i_m\}$ during the uplifting procedure.

Note that this tensor is weighted (although still non-negative) by construction. An analogous construction could be considered for weighted hypergraphs, although we have omitted it for the sake of clarity.

Continuing with the example hypergraph considered in Figure 1, we would have

$$T_{\sigma(123)} = 1, T_{\sigma(24)} = 1, T_{\sigma(35)} = 1 \xrightarrow{\text{Uplift}} T_{\sigma(123)} = 1, T_{\sigma(24\star)} = 1/3, T_{\sigma(35\star)} = 1/3, \qquad (3.3)$$

where $\sigma(ij) \in \mathfrak{S}_2$ and $\sigma(ijk) \in \mathfrak{S}_3$ denote any possible permutation of the indices. This is sensible because, for instance, previously there were two components describing the hyperedge $\{2, 4\}$, and now there are 6 describing $\{2, 4, \star\}$ but with diminished weight.

We want to stress the importance of this weighting choice, which sets the uplift apart from [25, 24]: Without it, we would not be able to claim that it appropriately generalizes the uniform HEC, as it would introduce spurious connectivity strengths. This ensures that the total adjacency relations are preserved. In the two cited works the actual strength of the connections is not considered, only the connectivity structure.

If the original hypergraph is strongly connected, the addition of a node to some hyperedges will not hinder the connectivity of the resulting hypergraph. Therefore, in that case it is simple to conclude the strong connectedness of the resulting (uplifted) hypergraph.

Lemma 3.4 (Strong connectedness of the uplifted hypergraph). *Let $H = (V, E)$ be a strongly connected hypergraph whose maximum hyperedge size is M and let $m > M$. Then, the uplifted hypergraph $\widetilde{H}$ is strongly connected.*

It is worth noting that, if the original hypergraph is disconnected, then the uplift may connect it. This enhanced connectivity is an artifact of the operation, and even though one will get a unique, well-defined centrality for that hypergraph, it might be of interest to carefully check whether that is preferable in the application/system in mind.

An uplifted nuance: Given that the uplift produces uniform hypergraphs, one could now think of applying either HEC or ZEC to the uplifted hypergraph, in order to obtain the centrality score of each node c_i. It turns out that there is a nuance in the uplifting procedure which impedes its usage in the ZEC case, which is not an issue in the HEC case, as we will now see.

Take, for example, a hypergraph H with only size 2 (pairwise) and 3 (triple) interactions. Suppose we uplift it adding an auxiliary node $\star$ once inside each of the pairwise edges. The aim of this procedure boils down to enabling the following rewriting of the "tensor apply" (present in both HEC 2.9 and ZEC 2.8) operation, grouping all the interactions together:

$$\sum_{j=1}^{n} a_{ij}^{(2)} c_j + \sum_{j,k=1}^{n} T_{ijk}^{(3)} c_j c_k \rightarrow \sum_{j,k=1}^{n} \widetilde{T}_{ijk} c_j c_k + \sum_{k=1}^{n} \widetilde{T}_{i\star k} c_\star c_k + \sum_{j=1}^{n} \widetilde{T}_{ij\star} c_j c_\star = \widetilde{T} \mathbf{c} \, \mathbf{c}. \qquad (3.4)$$

However, this involves a sum (the pairwise one) where the centrality of $c_\star$ would be involved. It would thus seem that there is a flaw in using the uplift to obtain centralities of the original hypergraph: They would depend on the centrality of the spurious node $\star$. Luckily, in the HEC centrality this is not the case, as we will show that this is an artifact of the procedure which we can get rid of.

To see this, consider applying HEC to an uplifted hypergraph $\widetilde{H}$, obtaining a centrality vector $\mathbf{c} = (c_1, ..., c_n, c_\star)^T$. As discussed before, if $\widetilde{H}$ is strongly connected, $\mathbf{c}$ is positive and unique up to scaling [22]. Therefore, one can always rescale it such that $c_\star = 1$. This choice solves the previously mentioned apparent contradiction, as the sums before and after the uplift would now coincide.

Moreover, given that the centrality scores we care about are just those of the "real" nodes, one would then just keep the centrality components associated with them, which can then be rescaled again at will (for instance, in order to normalize them). Hence, the initial scaling to achieve $c_\star$ was just a formal consistency check, but it can conveniently be ignored once we have computed the HEC solution.

The ZEC problem: Notice that the reason why we could ignore the aforementioned issue in the HEC case is the fact that if $\mathbf{c}$ is an $\mathcal{H}$-eigenvector, then $\mathbf{c}' \propto \mathbf{c}$ is still a $\mathcal{H}$-eigenvector. This is not the case for $\mathcal{Z}$-eigenvectors: They can not be rescaled and still solve the $\mathcal{Z}$-eigenproblem defined in Subsection 2.1 (recall that they are subject to a normalization constraint [8]).

It would seem that there is no use for the combination of uplift and $\mathcal{Z}$-eigenvectors. That turns out not to be the case, if we uplift an already 2-uniform hypergraph to a $(2+m)$-uniform hypergraph, as we will see. From the point of view of computing importance scores this is unnecessary (the ZEC could already be computed in the original hypergraph), but we will see that it plays an important role in the characterization of Perron-like $\mathcal{Z}$-eigenvectors for certain types of hypergraphs. This result will, in turn, feed back into the ZEC centrality quite naturally.

Therefore, from now on we will separate the discussion in two parts: on the one hand, if one starts with a non-uniform hypergraph, its uplift can be used to compute HEC-like centralities. On the other hand, if one starts with a uniform hypergraph, its uplift can shed light on properties of certain $\mathcal{Z}$-eigenproblems.

4. Uplift + $\mathcal{H}$-eigenvectors: Spectral centrality in non-uniform hypergraphs

We will now particularize what we have been discussing to the case of $\mathcal{H}$-eigenvectors. As we mentioned, the main interest of this uplift is the extension of the HEC centrality measure to the case of non-uniform hypergraphs. Given what we know so far, we can already do so.

Definition 4.1 (m-UHEC). *Let $H = (V, E)$ be a strongly connected hypergraph whose maximum hyperedge size is M and let $m \geq M$. The m-Uplifted $\mathcal{H}$-Eigenvector Centrality (m-UHEC) of the hypergraph H consists of the $n = |V|$ components associated with nodes in V of the HEC of the uplift of H to order m.*

For the sake of conciseness and to avoid cluttering the notation, we will from now on refer to the m-UHEC as just the UHEC, with the order being clear by the context, or specified otherwise.

Note that if H is already M-uniform and $m = M$, then the UHEC and standard HEC vectors coincide. It is straightforward to see that this measure is well-defined in the sense that the UHEC vector is positive and unique (up to scaling), as was the HEC measure.

Theorem 4.2 (Existence and uniqueness of the UHEC). *Under the assumptions of Definition 4.1, the UHEC vector exists and it is unique.*

Proof. Lemma 3.4 guarantees the strong connectedness of the uplifted hypergraph, and the Perron-Frobenius theorem for strongly connected hypergraphs [22] guarantees the existence and uniqueness of its HEC. □

The rest of this Section is organized as follows: In Subsection 4.1, we will apply the uplift operation to a pairwise graph, turning it into a 3-uniform hypergraph. This will show that, even though the HEC

of the uplifted graph is similar in ranking to the original eigenvector centrality, the reality is a bit more nuanced. In Subsection 4.2, we will supplement the uplifting of lower orders with a "projection" of higher ones, allowing us to define, not just one, but several uniformized centrality measures, one per order. A toy model exemplifies the advantages with respect to the vector centrality [23]. In Subsection 4.3, we will compare these methods with the approaches in [7, 26] over real and synthetic network data.

4.1. The pairwise case

The uplift procedure is not restricted to higher-order networks: One can also apply it to pairwise interaction networks. While in real applications there is no obvious reason why one would prefer it to other, well-established, spectral centrality measures, for us it will be interesting to discuss it as a means of comparison with them: If the centrality outcome after the uplift into a hypergraph was considerably different from the centrality outcome of the pairwise eigenvector centrality, that would signal a flaw in our approach.

Consider a pairwise interaction graph $G = (V, E)$ with adjacency matrix $A = (a_{ij})$. Uplift it to $\widetilde{H}^3 = (\widetilde{N}, \widetilde{E})$. Its 3-uniform tensor can be decomposed as

$$\widetilde{T}_{i_1 i_2 i_3} = \begin{cases} a_{i_1 i_2}/3 & \text{if } i_3 = \star, \\ a_{i_2 i_3}/3 & \text{if } i_1 = \star, \\ a_{i_1 i_3}/3 & \text{if } i_2 = \star, \\ 0 & \text{otherwise.} \end{cases} \tag{4.1}$$

With this decomposition, one can rewrite the HEC equation 2.9 as

$$\lambda c_i^2 = \sum_{j,k=1}^{n} T_{ijk} c_j c_k + \sum_{j=1}^{n} T_{ij\star} c_j c_\star + T_{i\star\star} c_\star c_\star = \frac{1}{3} \sum_{j=1}^{n} a_{ij} c_j c_\star, \tag{4.2}$$

$$\lambda c_\star^2 = \sum_{j,k=1}^{n} T_{\star jk} c_j c_k + \sum_{j=1}^{n} T_{\star j\star} c_j c_\star + T_{\star\star\star} c_\star c_\star = \frac{1}{3} \sum_{j,k=1}^{n} a_{jk} c_j c_k. \tag{4.3}$$

The second of these equations is not actually relevant: it just ties the value of the auxiliary node based on the scores of the rest. Requiring now the centrality of the auxiliary node to be $c_\star = 1$ we end up with

$$\lambda c_i^2 = \frac{1}{3} \sum_{j=1}^{n} a_{ij} c_j, \tag{4.4}$$

which resembles, up to the c_i^2, a weighted, undirected version of the Eigenvector Centrality Eq (2.1). It is therefore natural to compare the centralities obtained through this uplifted measure to those of the standard (pairwise) eigenvector centrality.

We expect to obtain a similar ranking (in the sense of the ordering of nodes by importance), although with a lower spread in the actual centrality scores. This is because, loosely speaking, the uplift compresses the centrality scores: The auxiliary node ties every node together, homogenizing the centrality. The most notable thing is, however, the fact that this homogenization can change the actual ranking between the nodes, as can be seen in Figure 2.
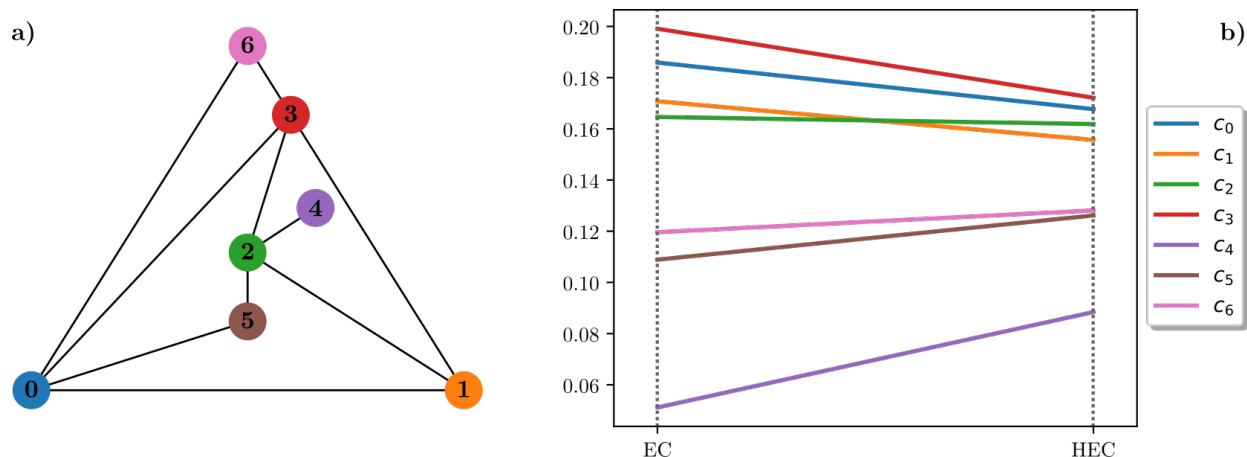
Figure 2. Panel **a)** shows the pairwise graph $G = (V, E)$ with 7 nodes whose Eigenvector Centrality (EC) and the HEC of its uplift to order 3 (HEC) are calculated. Panel **b)** shows the centrality values associated with the ℓ_1-normalized centralities in a parallel coordinate plot. The effect of the homogenization can be clearly seen, as well as the crossing in ranking between nodes 1 and 2.

4.2. Uplifting and projecting hypergraphs

So far we have discussed the way to deal with the hyperedges of size lower than the desired one, by means of uplifting those below it. But in order to truly embrace non-uniform hypergraphs we should also consider an operation bringing hyperedges of higher-orders down to the desired one.

The key to this has been hinted at when discussing the Clique motif Eigenvector Centrality, back in Subsection 2.1. There, an order-k hyperedge is split into all possible pairwise relations, $C(k, 2) = \binom{k}{2}$ of them, between its constituents. In other words, size k edges are projected into sets of size 2 edges. We can think of an analogous process but turning size k edges into sets of size $p < k$ edges.

Definition 4.3 (Projected hypergraph). *Let $H = (V, E)$ be a hypergraph whose maximum hyperedge size is M and let $2 \leq p < M$. Denote the set of hyperedges of size greater than p as E' and denote the set of all p-subsets of every element of E' as S. We define the projected hypergraph at order p as*

$$\widehat{H} = (V, \widehat{E}), \quad \widehat{E} = (E \backslash E') \cup S. \tag{4.5}$$

In other words, we can break apart each hyperedge of dimension k into $C(k, p) = \binom{k}{p}$ distinct hyperedges of dimension p.

Notice that, unlike what we did in the uplift case, we can not as of yet define an associated adjacency tensor, as $\widehat{H}$ will generally still be non-uniform. However, given that this operation entails, essentially, a substitution of each higher size edge by a collection of smaller ones, we need to discuss how to assign weights to the smaller ones generated from the projection.

If we follow a similar reasoning to the combinatorial one used in the uplift case (see Definition 3.3), one ends up with nonsensical weight assignments. Particularly it can be calculated to be

$$w = \frac{k!}{p! \, C(k, p)} = (k - p)!.$$

For instance, an order 4 hyperedge projected would be projected into order 2 hyperedges with weight $w = 2$. Hence the resulting hyperedges would have a higher participation than those already at the chosen order.

Instead, we can go back to Benson's work [7], and in particular the CEC calculation, which achieves a sensible projection assigning weights which are the result of counting how many times a pair participates in higher size edges. Our projection aims to generalize this concept, and thus the weights come from a similar counting argument (a p-subset's weight will be the number of higher-than-p order edges where the subset participates).

Joining everything together: As we mentioned before, the resulting projected hypergraph $\widehat{H}$ might not be uniform, and moreover, if we discard orders lower than p we are losing information, as if we uplift hyperedges discarding even higher interactions. For that reason, the key idea in terms of computing centralities is combining both projection of orders higher than a chosen order p and uplifting the lower ones.

Definition 4.4 (p-UPHEC). *Let $H = (V, E)$ be a hypergraph whose maximum hyperedge size is M and let $2 \leq p \leq M$. The p-Uplifted-Projected $\mathcal{H}$-eigenvector centrality (p-UPHEC) is the only positive $\mathcal{H}$-eigenvector of the uniform, weighted hypergraph resulting from*

(1) Adding an auxiliary node (or more than one as long as they are indistinguishable) to each hyperedge of size $k < p$ and weighting them with their corresponding combinatorial factor.

(2) Projecting down each hyperedge of size $k > p$ into a set of size p hyperedges, with their corresponding combinatorial factors.

As in the UHEC case, for the sake of conciseness we will from now on refer to the p-UPHEC as just the UPHEC, where again the order will be clear by the context, or specified otherwise.

It is straightforward to check that the connectivity of the resulting hypergraph is unchanged.

Lemma 4.5 (Strong connectedness of the projected hypergraph). *Let $H = (V, E)$ be a strongly connected hypergraph and $2 \leq p \leq M$. The hypergraph resulting from uplifting and projecting as in Definition 4.4 is strongly connected.*

Once again, we can easily show the consistency of this measure, as was the case with the UHEC.

Theorem 4.6 (Existence and uniqueness of the UHEC). *Under the assumptions of Definition 4.1, the UHEC vector exists and it is unique.*

Proof. Lemmas 3.4 and 4.5 guarantee the strong connectedness of the hypergraph resulting from projecting and/or uplifting, and the Perron-Frobenius theorem for strongly connected hypergraphs [22] guarantees the existence and uniqueness of its HEC. □

Note that there may be different UPHEC solutions associated with different values of the parameter p. To see this, consider the example from Figure 3.
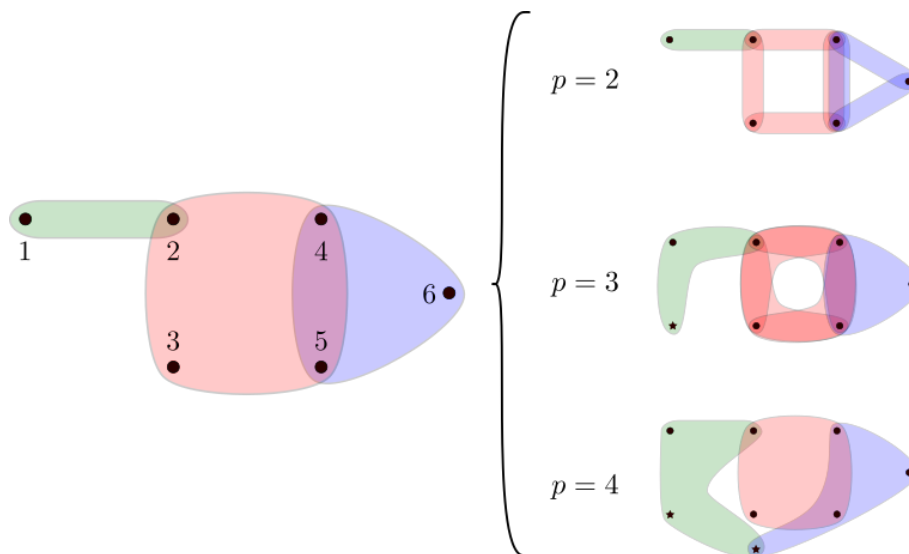
Figure 3. Example hypergraph and its three possible uniformizations at $p = 2, 3, 4$.

Example 4.7. *Let $H = (V, E)$ with $E = \{\{1, 2\}, \{2, 3, 4, 5\}, \{4, 5, 6\}\}$, and hence $M = 4$. There are three possible UPHEC vectors one can obtain, one for each $p = 2, 3, 4$.*

- *Case $p = 2$: This is equivalent to only considering the projection to order 2.*

$$H' = (V, E'), \quad \text{with} \quad E' = \{\{1, 2\}, \{2, 3\}, \{2, 4\}, \{2, 5\}, \{3, 4\}, \{3, 5\}, \{4, 5\}, \{4, 6\}, \{5, 6\}\}. \quad (4.6)$$

- *Case $p = 3$: In this case we mix the projection of the second hyperedge and the uplift of the first one, therefore computing the HEC of*

$$H' = (V', E'), \quad \text{with} \quad E' = \{\{1, 2, \star\}, \{2, 3, 4\}, \{2, 3, 5\}, \{3, 4, 5\}, \{4, 5, 6\}\}. \quad (4.7)$$

- *Case $p = 4$: This is equivalent to only considering the uplift to order 4, i.e., computing the 4-UHEC of*

$$\widetilde{H} = (\widetilde{V}, \widetilde{E}), \quad \text{with} \quad \widetilde{E} = \{\{1, 2, \star, \star\}, \{2, 3, 4, 5\}, \{4, 5, 6, \star\}\}. \quad (4.8)$$

Constructing the respective adjacency tensors and computing their Perron-like $\mathcal{H}$-eigenvector, we get the normalized centrality scores of Table 1.

Table 1. Centrality scores for each UPHEC case. Cells highlighted in green show the most central nodes in each case, while cells highlighted in red show the least central nodes. A good consistency check is the fact that the centralities of nodes 4 and 5 are identical in both cases, as they are indistinguishable in the original hypergraph H.

Case	c_1	c_2	c_3	c_4	c_5	c_6
$p = 2$	**0.0929**	0.1802	0.1690	**0.2084**	**0.2084**	0.1412
$p = 3$	**0.0623**	0.1949	0.1943	**0.2060**	**0.2060**	0.1364
$p = 4$	**0.0853**	0.1959	0.1953	**0.1993**	**0.1993**	0.1250

We can compare this method with the vector centrality one [23], which also distinguishes centralities order by order. The resulting centralities can be found in Table 2.

Table 2. Centrality scores yielded by the vector centrality measure at each order.

Order	c_1	c_2	c_3	c_4	c_5	c_6
2	0.5	0.5	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.3333	0.3333	0.3333
4	0.0	0.25	0.25	0.25	0.25	0.0

Here we see an example where the new measure improves upon existing ones, as it performs a similar task but it is capable of aggregating information of the whole hypergraph structure into each of the evaluated orders, rather than dismissing those which the nodes do not belong to.

In fact, one can see that once the whole structure is taken into account, in this example there would be no doubt about which is the least important node in the whole network and which are two of the most important ones. If one were to trust the vector centrality at second order, for instance, one could have been deceived into thinking that the first node is of rather remarkable importance. Moreover, the naïve way to combine these orders (summing the scores of each node) would also lead us to think that node 1 is more important than node 3, for example. It should be clear by now that the non-linear treatment is offering us valuable insights.

Notes on computational complexity: Before moving to real-world applications, we first want to address the computational cost of the algorithms discussed so far.

First, we need to discuss the creation of the tensor, which will have a different complexity depending on whether we are uplifting or projecting a hyperedge. In the case of the uplift, for every hyperedge that has to be uplifted we add the phantom node the necessary number of times (linear operation). It gets more complicated in the case of projecting, where we need to compute all the possible combinations of a hyperedge (factorial operation). Let $H = (V, E)$ be a hypergraph, $m \in \mathbb{N}$ the order we want to transform it to. Let $|E| = |E_u| + |E_m| + |E_p|$, where $|E_u|$ is the number of hyperedges that have to be uplifted, $|E_m|$ is the number of hyperedges already at the desired order, and $|E_p|$ is the number of hyperedges that have to be projected. Then, the overall number of operations that have to be done to create this weighted tensor is

$$\sum_{e_u \in E_u} (m - |e_u|) + \sum_{e_p \in E_p} m\binom{|e_p|}{m} = |E_u|(m - \bar{e}_u) + \sum_{e_p \in E_p} m\binom{|e_p|}{m}, \tag{4.9}$$

with $\bar{e}_u$ being the average size of the hyperedges that have to be uplifted. To compute the Big-O notation we have to choose the worst-case scenario, the highest-order term. In this case, it will be that associated with the projected edges

$$O\left(|E| \cdot m \cdot \binom{|e|}{m}\right). \tag{4.10}$$

Once we have created the tensor, we now need to compute the eigenvector corresponding to the largest $\mathcal{H}$-eigenvalue. In order to compute UHEC and UPHEC centralities, instead of creating a new algorithm, we used a variant of the power method with a weighted tensor (see [13]).

4.3. Numerical comparisons

A first attempt to generalize adjacency tensors in a non-uniform context was provided by [26] (and later glossed over by Benson in [7]), which goes by the name hyperedge "blowups" [20]. This method relies on suitably duplicating indices in the adjacency tensor to accommodate to higher-order hyperedges, and it has recently been computationally improved so as to avoid its high computational cost when it comes to the tensor apply operation [20]. However, as [8] already points out, there is some indeterminacy in this approach.

We will nevertheless consider the original (and only) proposal [26] which, given a hyperedge $e = \{v_1 \dots v_s\}$ with $2 \le s \le m$ nodes (where m is the maximum cardinality of the hyperedges), assigns it the m-uniform adjacency tensor components

$$a_{i_1 \dots i_m} = \frac{s}{\alpha} \quad \text{where} \quad \alpha = \sum_{p_1, \dots p_r = 1}^{n} \frac{m!}{p_1! p_2! \dots p_s!}, \tag{4.11}$$

and $i_1, \dots, i_m$ are chosen in all possible ways from $\{v_1, \dots, v_r\}$. The construction of this tensor is already disadvantageous. The time complexity of this uplifting method can be directly found by the intuitive idea behind it. Suppose we want to uplift the hyperedge e to order m. To do so, we will need all the possible combinations of adding each node to it, until we reach the desired order. Increasing the order by 1 would take $|e|$ operations (add each node to the hyperedge once). Increasing the order by 2, we would need to do $|e^2|$ operations (the operations mentioned before, and for each new hyperedge constructed, add each of the original nodes). It is straightforward to see that the time complexity we are talking about is $O\left(|e|^{m-|e|}\right)$ for each hyperedge to be uplifted. Nevertheless, this time complexity can be reduced through dynamical optimization to $O\left(|e|(m - |e|)\log m\right)$. The method proposed in this paper to uplift a hyperedge involves far fewer operations, having $O(m - |e|)$ for each hyperedge, as the only thing it is being done is adding a new node as many times as it is necessary. Moreover, this alternative uniformization does not include a notion of projection, which is why we have to supplement it with ours if one is interested in checking intermediate orders.

We now want to give a taste of the difference between the different tensorial methods discussed throughout this manuscript, namely: The standard HEC (Eq (2.9)), the UPHEC (Definition 4.4), and the alternative uniformization method (Eq (4.11)), at each of the different orders present in a hypergraph[‡].

4.3.1. Real-world hypergraph datasets

As further evidence of the interest and usefulness of the proposed method, some real-world hypergraph datasets are taken now into consideration, by analyzing three different points: How do the two hypergraph uniformizations discussed (our UPHEC and blowups [26, 20]) compare against the original, order-by-order analysis of the hypergraph put forward by Benson [7]? What is the difference

---

[‡]We will not be discussing the ZEC here, as it was already done in [7] and it does not have an UPHEC analogue, as discussed throughout the text.

between both uniformizations in these real cases. Lastly, even within either of these uniformizations, one needs to choose at which order to perform the analysis (uplifting lower orders and projecting higher ones in our method, "blowing up" lower orders and also projecting higher ones in the blowup one.

It is important to note that the besides the figures shown in the present manuscript, which have been picked due to their clarity and aid in the exposition, we perform a wider analysis of more hypergraphs (all of them freely available within the XGI library [27]), which can be found in the open repository at `https://github.com/LaComarca-Lab/non-uniform-hypergraphs`.

To start things off, let us consider two hypergraphs: A quintessential one, the Tags Ask Ubuntu dataset [28], also used in [7] to showcase the CEC, ZEC, and HEC proposals; and the Hypertext Conference one [29]. The former contains information about interactions within the *Ask Ubuntu StackOverflow* forum. Specifically, it can be seen as a hypergraph where nodes represent tags and hyperedges between tags represent questions marked with those tags. The latter contains data gathered during the *ACM Hypertext 2009* conference, pertaining to the interactions between its participants.

Some basic statistics of these hypergraphs (after pre-processing them with the XGI library [27] in order to remove isolated nodes, singleton edges, etc.) can be observed in Table 3. Note that when studying each uniform order as isolated some nodes will become disconnected if they have no such interactions.

Table 3. Number of nodes and hyperedges at each order of the hypergraphs, as well as the size of the Largest Connected Component (LCC) containing them compared to the total. There is a small discrepancy between the sum of hyperedges at each order and the total number, due to the fact that some hyperedges become disconnected if we remove other orders and therefore they will not appear in the LCC.

	tags_ask_ubuntu [28]			hypertext-conference [29]		
Order	Nodes	Hyperedges	Size of LCC	Nodes	Hyperedges	Size of LCC
2	2714	28134	89.84%	113	2103	100%
3	2821	52282	93.38%	105	302	92.92%
4	2722	39158	90.10%	11	12	9.73%
5	2564	25475	84.87%	8	7	7.08%
6	-	-	-	8	4	7.08%
Complete	3021	145053	100%	113	2434	100%

The natural way to compare rankings is by means of some correlation measure which only takes into account the ordinal correlation between the entries (i.e., their position within the ranking) rather than their actual magnitudes. One of the best-known examples of this measure is Kendall's tau correlation coefficient ($\tau \in [0, 1]$, where the closer to 1 the more correlated), which we will compute between every pair of rankings.

Before showing the actual results, we should mention that in order to compare two rankings, they must contain the same number of elements. However in the uniformized versus non-uniformized cases

this is not the case (the non-uniformized, i.e., standard HEC versions only keep the LCC with those interactions). For that reason we have chosen to fill the empty entries with a zero value, as they do not participate in such an order. It is here that we can already glimpse at the issue with the standard, non-uniformized HEC: If we look at Table 3 we can see that, while this may be sensible in the `tags_ask_ubuntu` dataset, in the `hypertext-conference` the LCC of order beyond 3 is a minuscule part of the entire hypergraph. For this reason, the ranking will be localized around those nodes, yielding $\tau \approx 0$.

The results of the comparison between each of the rankings are shown in Figure 4. At this stage, we will focus on the first of the questions described before: The comparison of either uniformization with the non-uniform approach per order.
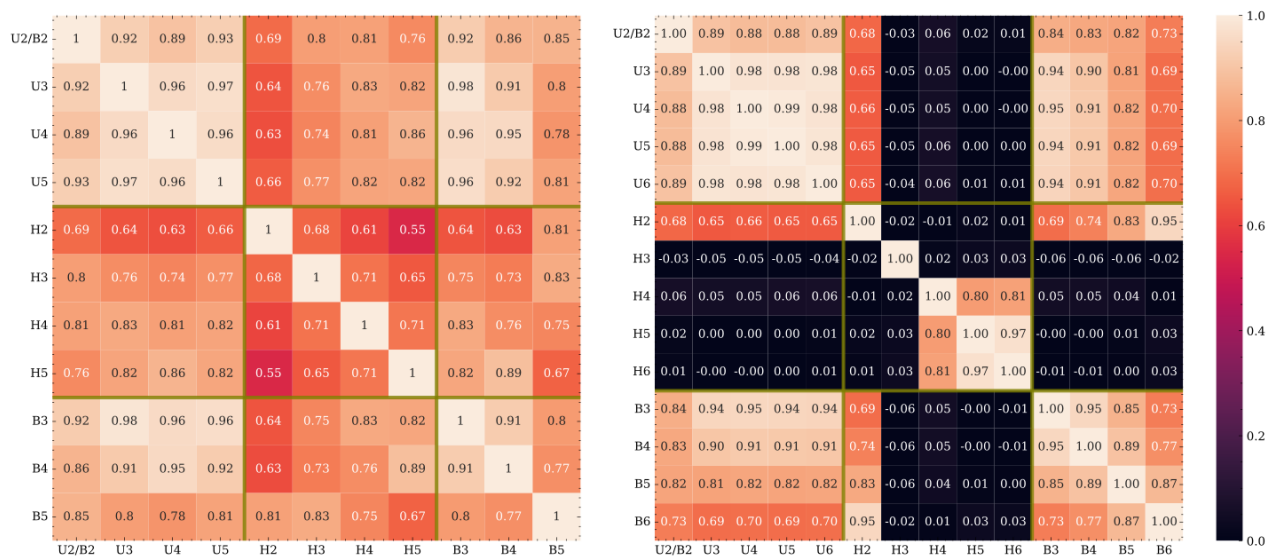

Figure 4. Kendall's tau correlation coefficient between the whole rankings obtained in each of the methods, for the `tags_ask_ubuntu` (left) and `hypertext-conference` (right) datasets. Methods are labeled as U2, U3, U4, and U5 for the UPHEC case; H2, H3, H4, and H5 for the HEC at each order, and B3, B4, and B5 for the blowup uniformization discussed above. U2 and B2 are equal as they only include the projection, and are therefore shown together.

In the `tags_ask_ubuntu` dataset, the four standard HEC measures among themselves have the lowest correlations. The lowest correlation of the whole figure is actually that between the 2nd and 5th orders. This is product of the fact that the uniform hypergraphs at each order have little to do with each other, they each describe a portion of the whole.

As we advanced before, this is much more drastic in the `hypertext-conference` dataset: There almost every correlation yields a number close to zero, except for the one between orders 5 and 6, as the LCC's of these orders share the same 8 nodes (see Table 3).

This analysis makes it clear that there is a need for uniformized versions of the HEC centrality, as the order-by-order study of hypergraphs clearly lacks a cohesive description of the whole[§].

[§]For a visual analogy of what is going on, check the cover of the first edition of "Gödel, Escher, Bach: An Eternal Golden Braid" by Douglas R. Hofstadter [30].

Having dealt with the question of uniform measures versus non-uniform ones, we now shift our focus to the other two problems: the comparison between uplift and blowups, and the order to inspect them in. In order to get a better understanding, we supplement the previous examples with other four real hypergraph datasets, also available in XGI, whose most basic statistics (this time without an order-by-order overview) after preprocessing are summarized in Table 4.

Table 4. Number of nodes, hyperedges, and maximum order of each hypergraph (after removing isolated nodes, and duplicated edges, and keeping the LCC), as well as the average Kendall-tau coefficient between the UPHECs $\langle \tau_{UU} \rangle$ and the blowups $\langle \tau_{BB} \rangle$ at each order between 3 and the respective maximum order.

Dataset	Nodes	Hyperedges	Maximum order	$\langle \tau_{UU} \rangle$	$\langle \tau_{BB} \rangle$
`tags_ask_ubuntu` [28]	3021	145053	5	0.960	0.825
`hypertext-conference` [29]	113	2434	6	0.982	0.844
`contact-primary-school` [31]	242	12704	5	0.962	0.905
`contact-high-school` [32]	327	7818	5	0.946	0.863
`sfhh-conference` [33]	403	10541	9	0.918	0.748
`diseasome` [34]	516	314	11	0.724	0.590

For each of these hypergraphs, the correlation between the UPHEC and blowup+projection measures have been computed in Figure 5, now ignoring the non-uniform measures for ease of visualization[¶], as well as the top-most column and left-most row (the "U2/B2" ones in Figure 4), for they correspond to only projecting any higher-order interaction to pairwise ones, and computing the HEC of the corresponding, uniform hypergraph. As there are no uplifts nor blowups, there is no distinction on the uniformization used, which is why we choose to ignore them at this point.

We can draw the following conclusions from Figures 4 and 5, with respect to both uniformization procedures.

First, the average correlation between the same type of uniformizations at different orders (i.e., the U-U and B-B quadrants) is always higher in the uplift than in the blowup cases. For reference, the average value of off-diagonal correlations in those two quadrants, for each hypergraph, are shown in Table 4. In that regard, it is interesting to look at the `sfhh-conference` example: The lowest correlation between any two UPHEC measures is found to be around 0.88, while the lowest in the blowup uniformization is around 0.55.

Moreover, one can clearly see that the higher the order inspected, the more disparity. This is pointing to the fact that, while in lower orders the projection part of the measure (which is the same in both the UPHEC and in the blowup uniformization) is evening the rankings, when we focus on the highest order (thus only having vanilla uplift and blowup, no projection) the blowup is computing something slightly different. In that sense, this seems to confirm the claim in [8] about the blowup uniformization and the fact that it contains a degree of arbitrariness in the augmentation, something which we indeed observe.

[¶]In the `diseasome` case we have also restricted the computation to orders up to 9 for better visualization.
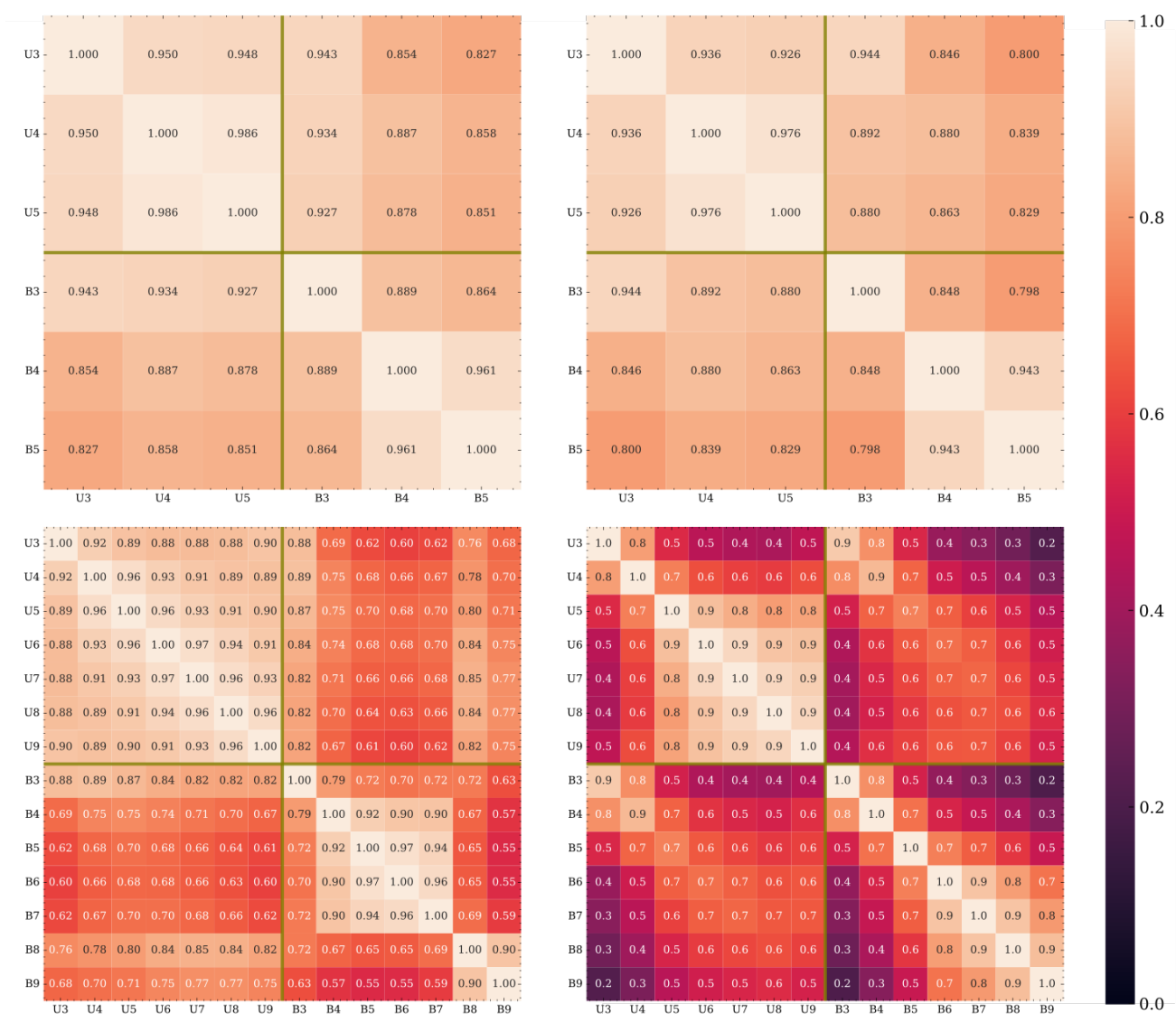
Figure 5. Kendall's tau correlation coefficient between the rankings obtained by the UPHEC and blowup uniformizations (with the same labeling conventions as in Figure 4) for the `contact-primary-school` (top-left), `contact-high-school` (top-right), `sfhh-conference` (bottom-left), and `diseasome` (bottom-right) datasets.

Apart from the choice of uniformization procedure, we wanted to understand the implications of the choice of order at which to inspect the hypergraph. Focusing on the UPHEC method, what we can see is that in most examples the choice is basically irrelevant: Once we take into account every level of interaction (either through projection or uplift), a centrality unison emerges, something we can clearly see from $\langle \tau_{UU} \rangle \approx 1$. Nevertheless, the correlation is better at higher-orders, meaning that the more uplift and less projection, the more agreement in the description of the overall hypergraph.

At this point it is important to also consider the computational cost of each of these methods. As we have discussed, ideally one would want to compute the centrality with UPHEC at the highest order available. However, it might be preferable to have a balance between uplift and projection, staying therefore at intermediate orders. Alternatively, and if computational efficiency is a necessity, one could

use the method proposed in [20], which achieves a remarkable speed increase in the computation of the blowup, turning an $O(n^M)$ problem into one being polynomial in M, the maximum order.

Aside from the full ranking comparison, it is often interesting to understand how the correlation changes when we contrast the top K nodes obtained with one method with their corresponding ranking obtained according to another method, as we increase the amount K of nodes sampled. For the sake of simplicity we will show this with the `tags_ask_ubuntu` case, as the results are similar in the rest of the datasets.

Given the amount of possible comparisons (12 in the case of UPHEC-UPHEC, 16 in the cases of UPHEC-HEC, etc.), we have decided to filter out most of them in order to present a meaningful figure. In particular, for each measure comparison we have chosen to keep at most correlations: the correlation reaching the highest maximum, the correlation reaching the lowest minimum, and the two correlations whose average is minimum and maximum. We feel that these conditions will provide us with a set of correlations which can convey more information (in the sense of the most similar and dissimilar rankings). The resulting plot is displayed in Figure 6, and the unfiltered one can be found in the open repository available at `https://github.com/LaComarca-Lab/non-uniform-hypergraphs`.
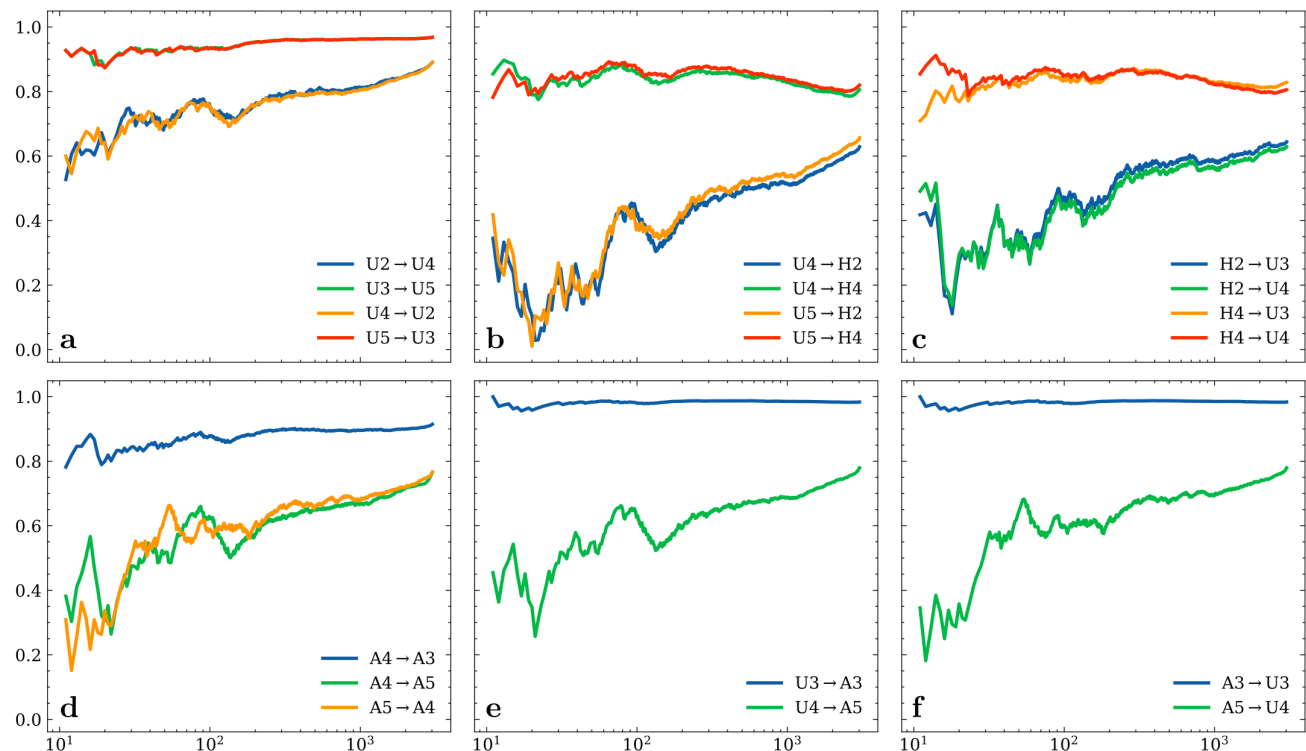

Figure 6. Kendall's tau correlation coefficient (log scale) between the top K ranked nodes from one method with the same nodes from another method on the `tags-ask-ubuntu` co-tagging dataset. Panels **a**, **b**, **c**, **d**, **e**, and **f** represent UPHEC-UPHEC, UPHEC-HEC, HEC-UPHEC, Alternative-Alternative, UPHEC-Alternative, and Alternative-UPHEC, respectively.

We see that despite some initial fluctuations around $K = 100$, most correlations tend to increase or stabilize, converging to their respective values shown in Figure 4. We also notice that in most cases the

minimums are reached in pairs, e.g., U2 and U4 are not very correlated with each other in Subfigure 1**a** in either direction, which is rather sensible.

4.3.2. Synthetic hypergraphs

In this subsection, we introduce some experiments that will help confirming the robustness of the proposed method. To do so, we use several synthetic hypergraphs, not only as a source of information, but also to prove that the new proposed measure works as expected in any kind of hypergraph, regardless of the domain to which the hypergraph belongs.

Synthetic hypergraphs were generated using the method proposed in [35], which extends the binomial Erdős-Rényi random graph model to non-uniform hypergraphs. This method was chosen because of the extension of the properties of the Erdős-Rényi model to non-uniform hypergraphs. Furthermore, it is particularly well suited to our specific requirements: Given a tuple of probabilities, generate a hypergraph where each hyperedge of size n has the corresponding probability in the tuple, more on this later. This ensures that the hypergraph generated is non-uniform. Furthermore, this method is among the most computationally efficient within the XGI library [27], as it is one of the fastest algorithms for generating random non-uniform hypergraphs.

The method works as follows: In the pairwise case, this method works by choosing p_2, which expresses the probability of any two nodes v_1 and v_2 forming the edge (v_1, v_2). In the extension, to generate a *random non-uniform hypergraph* with n nodes and hyperedge sizes $\{2, ..., m\}$, we provide probabilities $(p_1, ..., p_m)$, where each p_i expresses the probability of forming i-hyperedges between any i nodes in the hypergraph, and generate them accordingly.

The generated hypergraphs used here have $n = 100$ nodes and the hyperedge sizes range from 2 to 5. We selected $p_2 = 0.1 > \log(n)/n$ such that we can ensure that the generated hypergraph is strongly connected and $p_5 = 10^{-6}$ such that each generated hypergraph has approximately 100 5-hyperedges.

By fixing these parameters, we iterate through p_3 and p_4, assuring that we would have around 150 and 200 different 3- and 4-hyperedges as the minimum, and around 1500 and 2000 as the maximum, respectively. For this purpose, both p_3 and p_4 were taken from two equally spaced ranges, each split into eight equally spaced parts with $p_3 \in (10^{-3}, 10^{-2})$ and $p_4 \in (5 \cdot 10^{-5}, 5 \cdot 10^{-4})$.

In this way, we have 64 possible combinations of (p_3, p_4) to generate hypergraphs. We generated 20 hypergraphs for each 4-tuple (p_2, p_3, p_4, p_5). For each of these 1280 hypergraphs, we computed the k-UPHEC with $k \in \{2, 3, 4, 5\}$ and computed Kendall's tau for all combinations of measures, as shown in Figure 7.

As we can see in Figure 7, the correlations can be split into two disjoint behaviors, the upper and lower rows. The upper row shows the proportional growth in the correlation with p_3. We discuss the effects of p_4 on the different images in this row. Having noted that both p_2 and p_5 are fixed, if we compare "U2 vs U3", the growth of p_3 is proportional to the growth of the correlation. One can easily identify that the reason behind this is that having more 3-hyperedges will compensate for the difference in size between 4- and 5-hyperedges and 2-hyperedges. In addition, having more 3-hyperedges makes the 3-UPHEC closer to the original hypergraph. The opposite occurs in the case where we introduce more 4-hyperedges; we will not only have the difference between the 2-hyperedges, but also with 3-hyperedges.
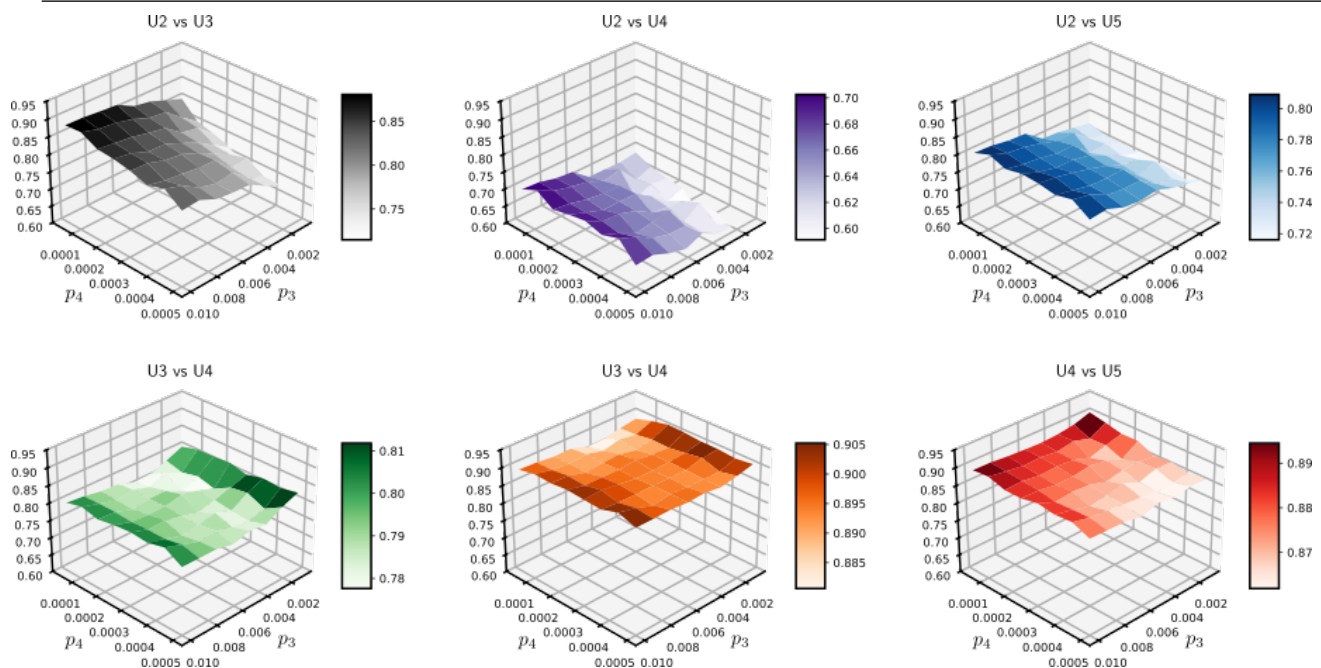
Figure 7. Average Kendall's tau correlation coefficient between the rankings obtained for each of the generated hypergraphs. "Ux vs Uy" represents the correlation between the UPHECs at orders x and y.

Note that in the case "U2 vs U4" we can only see a proportional growth with p_3, but not the inverse proportional growth with p_4 as much as before. It is trivial to compute the 4-UPHEC; introducing 4-hyperedges will have a positive effect, as it introduced 3-hyperedges in the previous case. Here, p_4 also compensates for the effect of 2- and 4-UPHEC on the 5-hyperedges, as does (and did before) p_3. Finally, if in the "U2 vs U5" case, the effect of p_4 appears to be irrelevant, we detect a strong effect of p_3, as 3-hyperedges are again acting as an intermediate between the 2- and 5-UPHEC.

Now that we have discussed the obtained computations on synthetic hypergraphs, we can conclude that regardless of the differences for different pairs (p_3, p_4), the proposed uniformization keeps track of the centrality measures, where each node has a similar ranking position in the different uniformizations; that is, the importance of each node is relatively preserved.

5. Uplift + $\mathcal{Z}$-eigenvectors: Uniqueness sufficient condition

As we previously discussed, the ZEC centrality can not be computed when uplifting a non-uniform hypergraph, as the different sums can only be grouped together if we are able to rescale the centrality such that $c_\star = 1$, which we can not if we are considering $\mathcal{Z}$-eigenvectors. However, this is not an issue if we start from a 2-uniform hypergraph (in other words, a pairwise graph).

In fact, $\mathcal{Z}$-eigenvectors allow us to generalize the uplift operation to having more than one different auxiliary node[‖], e.g., $\star_1, ..., \star_k$.

[‖]This generalization was already possible in the HEC case, however in that case it only cluttered the notation and hampered the calculation, as the computational complexity scales with the number of *distinct* nodes involved. Note also that in that case further conditions would be required for a well-defined uplift, as in order to be able to scale the centrality such that $c_{\star_k} = 1 \; \forall k$, we need all of them to be indistiguishable from each other, i.e., they must be related by permutation.

Definition 5.1 (Multi-Uplifted hypergraph). *Let $H = (V, E)$ be an M-uniform hypergraph and let $m \geq M$. We define the multi-uplifted hypergraph at order m with s auxiliary nodes, each contained p_k times within each hyperedge as*

$$\widetilde{H} = (\widetilde{V}, \widetilde{E}), \quad where \quad \widetilde{V} = V \cup \{\star_1, ..., \star_s\} \quad and \quad \widetilde{E} = \left\{ e \cup \left(\bigcup_{k=1}^{s} \bigcup_{l=1}^{p_k} \{\star_k\} \right), e \in E \right\}, \tag{5.1}$$

with $\sum_{k=1}^{s} p_k = m - M$.

As we previously declared, this operation on 2-uniform (standard) graphs allows us to relate the $\mathcal{Z}$-eigenvectors of the adjacency tensor of hypergraphs to those of their original, standard graph. To see this, consider adding two different auxiliary nodes $\star, \times$ to a graph G with adjacency matrix $A = (A_{ij})$. This operation translates into the following rewriting:

$$\sum_{j=1}^{n} A_{ij} c_j \longrightarrow \sum_{j,k,l=1}^{n, \star, \times} \widetilde{T}_{ijkl} c_j c_k c_l = \binom{3}{2} \sum_{k,l=1}^{n} \widetilde{T}_{ij\star\times} c_j c_\star c_\times, \tag{5.2}$$

where the notation $\sum_{i=1}^{n, \star, \times}$ indicates summing over the index $i \in [n] \cup \{\star, \times\}$.

Given that, by definition, $\widetilde{T}_{ij\star\times} = \frac{1}{12} A_{ij}$, then the $\mathcal{Z}$-eigenvector equation of the uplifted 4-uniform hypergraph is equivalent to the $\mathcal{Z}$-eigenvector equation of the original 2-uniform hypergraph, which reduces to the standard eigenvector centrality of the graph, i.e.,

$$\lambda \mathbf{c} = \widetilde{T} \mathbf{c}^3, \quad \mathbf{c} = (c_1, ..., c_n, c_\star, c_\times)^T \iff \lambda' \mathbf{c}' = A(\mathbf{c}')^2, \quad \lambda' = \frac{4\lambda}{c_\star c_\times}, \quad \mathbf{c}' = (c_1, ..., c_n)^T. \tag{5.3}$$

We can extrapolate this example to the uplift from a 2-uniform hypergraph to a $(2 + l)$-uniform hypergraph, as stated in the following result.

Theorem 5.2 (Correspondence between $\mathcal{Z}$-eigenvectors). *Let $\widetilde{H} = (\widetilde{V}, \widetilde{E})$ be a strongly connected, $(2 + l)$-uniform hypergraph with $l \geq 1$. If there is a non-empty subset of nodes $V^\star = \{\star_1, ... \star_s\} \subset \widetilde{V}$, each contained $\{p_1, ..., p_s\}$ times, respectively, in every hyperedge, such that $\sum_{i=1}^{s} p_i = l$, then,*

- *the components of the $\mathcal{Z}$-eigenvectors of $\widetilde{H}$ associated with the nodes $n \in V = \widetilde{V} \backslash V^\star$ correspond to those of the 2-uniform hypergraph $H = (V, E)$ having those s nodes removed;*
- *the components of the positive $\mathcal{Z}$-eigenvectors of $\widetilde{H}$ associated with the auxiliary nodes $n \in V^\star$ are uniquely determined by the other components;*
- *the $\mathcal{Z}$-eigenvalues $\widetilde{\lambda}$ of $\widetilde{H}$ correspond to those of the 2-uniform hypergraph H, λ, rescaled as*

$$\widetilde{\lambda} = \lambda \Omega (\widetilde{c}_{\star_1})^{p_1} ... (\widetilde{c}_{\star_s})^{p_s}, \quad \Omega = \frac{(l + 1)!}{\prod_{i=1}^{s} p_i!}. \tag{5.4}$$

Proof. Under the conditions stated, $\widetilde{H}$ can be viewed as an uplift of the hypergraph H with s auxiliary nodes, each one contained equally in each and every hyperedge. The $\mathcal{Z}$-eigenvector equation for the uplifted hypergraph can be written as

$$\widetilde{\lambda} \widetilde{c}_{i_1} = \sum_{i_2,...,i_{2+l}=1}^{n, \star_1,...,\star_s} \widetilde{T}_{i_1,...,i_{2+l}} \widetilde{c}_{i_2} ... \widetilde{c}_{i_{2+l}} = \Omega \sum_{i_2=1}^{n} T_{i_1,i_2} \widetilde{c}_{i_2} (\widetilde{c}_{\star_1})^{p_1} ... (\widetilde{c}_{\star_s})^{p_s}, \tag{5.5}$$

where we have summed over the auxiliary nodes, recovering the pre-uplifted tensor T components times a combinatorial factor Ω, product of the symmetry of the adjacency tensor. We now carefully calculate this factor.

(1) In the equation for node i, there will be a sum over $2 + l - 1 = l + 1$ indices (1 corresponds to its real j-th neighbor and l to the auxiliary nodes added). We will have all possible $(l + 1)!$ permutations.

(2) We need to subtract the repetitions of auxiliary nodes, given by their multiplicities p_i.

Having both of these facts considered we can easily calculate it to be

$$\Omega = \frac{(l + 1)!}{\prod_{i=1}^{s}(p_i!)}. \tag{5.6}$$

The centralities of these auxiliary nodes can now be pulled out of the sum, obtaining

$$\lambda \widetilde{c}_{i_1} = \sum_{i_2=1}^{n} T_{i_1,i_2} \widetilde{c}_{i_2}, \quad \widetilde{\lambda} = \lambda \Omega (\widetilde{c}_{\star_1})^{p_1} ... (\widetilde{c}_{\star_s})^{p_s}, \tag{5.7}$$

where we have already rescaled the $\mathcal{Z}$-eigenvalue accordingly. Noticing that $T = A$ with A being the adjacency matrix of G, we arrive at the equation $\lambda \mathbf{c} = A\mathbf{c}$, which is precisely the eigenvector equation of the 2-uniform hypergraph (pairwise graph) G.

Therefore, the first n components of the $\mathcal{Z}$-eigenvector of the uplifted hypergraph $\widetilde{H}$ correspond to the eigenvector $\mathbf{c}$ associated with G.

It is left for us to discuss the behavior of the remaining equations, one per auxiliary node. Without loss of generality, we consider the equation of node $\star_1$:

$$\widetilde{\lambda} \widetilde{c}_{\star_1} = \sum_{i_2,...,i_{2+l}=1}^{n, \star_1,...,\star_s} \widetilde{T}_{\star_1,...,i_{2+l}} \widetilde{c}_{i_2} ... \widetilde{c}_{i_{2+l}} = \Omega \sum_{i_1,i_2=1}^{n} T_{i_1,i_2} \widetilde{c}_{i_1} \widetilde{c}_{i_2} (\widetilde{c}_{\star_1})^{p_1-1} ... (\widetilde{c}_{\star_s})^{p_s}. \tag{5.8}$$

Multiplying both sides by $c_{\star_1}$ and then replacing $\widetilde{\lambda}$ in terms of λ, we obtain the following expression:

$$\lambda (\widetilde{c}_{\star_1})^2 = \sum_{i_1,i_2=1}^{n} T_{i_1,i_2} \widetilde{c}_{i_1} \widetilde{c}_{i_2}. \tag{5.9}$$

Therefore, each component associated with an auxiliary node is uniquely determined (up to a sign, although we can always choose the positive solution) by the components of the non-auxiliary nodes.

□

Remark 5.3. *We have omitted the norm constraint required for $\mathcal{Z}_1$- or $\mathcal{Z}_2$-eigenvectors. We are allowed to do so because we uplift a pairwise graph: Eigenvectors of the adjacency matrix can be rescaled at will, and therefore the first n components of the $\mathcal{Z}$-eigenvector of the uplifted hypergraph $\widetilde{H}$ can be matched to a specific scaling of the eigenvector of the adjacency matrix of G.*

Note that this is the reason why this theorem can not be generalized to an uplift from an m-uniform hypergraph to an $(m+l)$-uniform hypergraph: Even though the $\mathcal{Z}$-eigenvector equations can be related to each other, in general their norm constraints will be incompatible.

For illustrative purposes we provide an example which can be analytically solved, following the uplift on Figure 8.
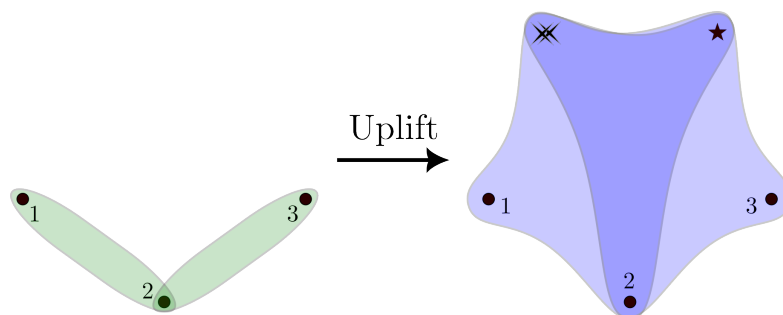

Figure 8. Uplift of a 2-uniform hypergraph (a pairwise graph) to 5-uniform by adding two nodes $\star, \times$ to each hyperedge, the latter being added twice (depicted as there being two of them for illustrative purposes).

Example 5.4. *Consider the graph $G = (V, E)$ with nodeset $V = \{1, 2, 3\}$ and edgeset $E = \{\{1, 2\}, \{2, 3\}\}$). It can be seen as a 2-uniform hypergraph $H \simeq G$. Suppose we uplift it to a 5-uniform hypergraph $\widetilde{H}$, adding auxiliary nodes $\star$ and $\times$ in each hyperedge, the former once and the latter two times, i.e.,*

$$\widetilde{H} = (\widetilde{V}, \widetilde{E}), \quad \widetilde{V} = \{1, 2, 3, \star, \times\}, \quad \widetilde{E} = \{\{1, 2, \star, \times, \times\}, \{2, 3, \star, \times, \times\}\}, \tag{5.10}$$

as shown in Figure 8.

The first thing we would need to do is rescale the adjacency matrix into the hypergraph tensor with suitable combinatorial factors (as in 3.3). However, here we can omit this step, as this factor is the same for all components. This implies that it will only modify the $\mathcal{Z}$-eigenvalue, but that is something we will already compute.

The $\mathcal{Z}$-eigenvector equation of $\widetilde{H}$ decouples into three distinct ones:

- *Three equations for the centrality of the original nodes ($i \in \{1, 2, 3\}$):*

$$\lambda c_i = \sum_{j,k,l,m} T_{ijklm} c_j c_k c_l c_m = \frac{4!}{2!} \sum_{j=1}^{n} T_{ij\star\times\times} c_j c_\star c_\times^2 = 12 c_\star c_\times^2 \sum_{j=1}^{n} A_{ij} c_j, \tag{5.11}$$

  where this combinatorial factor is the product of fixing 4 indices, out of which 2 are repeated.

- *An equation for the centrality of the auxiliary node $\star$:*

$$\lambda c_\star = \sum_{j,k,l,m} T_{\star jklm} c_j c_k c_l c_m = \frac{4!}{2!} (T_{\star 12\times\times} c_1 + T_{\star 23\times\times} c_3) c_2 c_\times^2$$
$$= 12 (A_{12} c_1 + A_{23} c_3) c_2 c_\times^2. \tag{5.12}$$

- *An equation for the centrality of the auxiliary node $\times$:*

$$\lambda c_\times = \sum_{j,k,l,m} T_{\times jklm} c_j c_k c_l c_m = 4! (T_{\times 12\star\times} c_1 + T_{\times 23\star\times} c_3) c_2 c_\star c_\times$$
$$= 24 (A_{12} c_1 + A_{23} c_3) c_2 c_\star c_\times. \tag{5.13}$$

If we rescale $\lambda' = \lambda/(12c_\star c_\times^2)$, we have that the first of them becomes $\lambda'\mathbf{c} = A\mathbf{c}$; in other words, it is the eigenvector equation of the adjacency matrix of the original graph G. As it is connected, we are guaranteed to have a unique, positive solution $\mathbf{c} > 0$.

The remaining equations are then (almost) completely fixed, as after rescaling the eigenvalue leads to

$$\lambda' c_\star^2 = (A_{12}c_1 + A_{23}c_3)\, c_2, \quad \lambda' c_\times^2 = 2(A_{12}c_1 + A_{23}c_3)\, c_2, \tag{5.14}$$

which not only enforces $\sqrt{2}c_\star = c_\times$ but also guarantees their positivity, as $A_{12} = A_{23}$ and $\mathbf{c} > 0$.

Yet, there is a subtlety to take into account: even though the Perron eigenvector $\mathbf{c}$ can be rescaled as $\mathbf{c}' = \alpha\mathbf{c}$, the $\mathcal{Z}$-eigenvector including $c_\star, c_\times$ cannot, as it requires some normalization ($\mathbf{c}^T\mathbf{c} = 1$ or $\|\mathbf{c}\|_1 = 1$), which will force upon your solution the suitable value of α.

With all these taken into account, we find the unique, positive solution $\tilde{c} = (\mathbf{c}, c_\star, c_\times)^T$ of the problem to be

$$\mathcal{Z}_1 : \tilde{c} = \frac{1}{4 + 2\sqrt{2}}\left(1,\ \sqrt{2}, 1,\ \sqrt{2}, 2\right)^T; \quad \mathcal{Z}_2 : \tilde{c} = \frac{1}{5}\left(1,\ \sqrt{2}, 1,\ \sqrt{2}, 2\right)^T. \tag{5.15}$$

We can finally obtain the following sufficient condition for existence and uniqueness of certain $\mathcal{Z}$-eigenvectors of tensors.

Corollary 5.5 (Sufficient condition for the existence of the Perron-like $\mathcal{Z}$-eigenvector). *Let T be a symmetric tensor of order m > 2. If its associated hypergraph H is strongly connected and can be seen as an uplift from a pairwise graph G, then a Perron-like $\mathcal{Z}$-eigenvector of T (i.e., a unique, positive $\mathcal{Z}$-eigenvector) is guaranteed to exist.*

Proof. This follows directly from the Perron-Frobenius theorem, as it guarantees the existence and uniqueness of the eigenvector $\mathbf{c}$ of the graph G and the fact that the remaining (auxiliary nodes) equations fix uniquely (after choosing their positive values) these components in terms of $\mathbf{c}$. □

The only thing left for us to discuss is the connection between this multilinear algebra result, and our original perspective, which was that of hypergraph centralities. But making this leap is rather evident.

Corollary 5.6 (Sufficient condition for the uniqueness of ZEC). *If a hypergraph H is strongly connected and can be seen as an uplift from a pairwise graph G, then it has a unique Perron-like $\mathcal{Z}$-eigenvector.*

It is important to point out that the computation of $\mathcal{Z}$-eigenvalues is particularly complicated (see, e.g., [36, 7, 19]), however our work clears the path for the simple computation of a whole class of hypergraphs.

6. Conclusions and future work

In this study, we introduced a novel approach to analyze non-uniform hypergraphs by transforming them into an uniform hypergraph with the addition of an auxiliary node and suitably adjusting the weights of the transformed hyperedges, in an operation which we refer to as the "uplift". This transformation enabled us to apply well-defined centrality measures based on the eigenvectors of the resulting adjacency tensor. Through extensive comparisons with existing centrality measures in the literature, we have demonstrated the efficacy and relevance of our approach.

The key contribution of this work lies in the ability to bridge the gap between non-uniform hypergraphs and well-established centrality metrics. By introducing the auxiliary node, we effectively translated complex, multifaceted relationships into a format that aligns with already established hypergraph analysis techniques based on the $\mathcal{H}$-eigenvector centrality, in a manner that is consistent due to the weighting choice. This, when supplemented with a projection operation, furnishes a sensible, well-defined novel centrality measure which, while at the same time retaining some degree of granularity (in the order which we can put the focus on), yields similar results, hence agreeing on the most important nodes of a hypergraph.

Our results showcased the advantages of our approach over existing methods: On the one hand the uniformization allows us to incorporate more information to the centrality when compared to uniform methods, and on the other hand computing the adjacency tensor has a much lower computational complexity than the single other method available in the literature. Moreover, from an algebraic point of view, we see that a generalization of the uplift to different nodes sheds light on the characterization of $\mathcal{Z}$-eigenvectors of tensors. In particular it provides a simple route to their computation for a particular class of hypergraphs.

This work paves the way for a new era in the analysis of higher-order systems, allowing for an identification of the most central nodes in cases where it was previously unfeasible or, at best, undecidable, in a way that is not very computationally demanding. Here we exemplified these methods with social and biological data, but data from other sources could be gathered to further validate these findings. Lastly, while we considered static undirected hypergraphs in this work, it would be interesting to generalize the uplift to other situations, like directed or heterogeneous hypergraphs [37, 38] or evolving ones [39].

In summary, our study has presented a promising framework for the analysis of non-uniform hypergraphs, making them amenable to well-defined centrality measures based on tensor eigenvectors. This advancement holds great potential for applications across various domains, including social networks, biological systems, transportation networks, and beyond. By providing a bridge between complex, non-uniform relationships and established network analysis techniques, our approach contributes to a deeper understanding of the underlying structures and the identification of critical nodes within these intricate systems.

Author contributions

G. Contreras-Aso and C. Pérez-Corral: Conceptualization, data curation, formal analysis, software, validation, visualization, writing – original draft, writing – review & editing; M. Romance: Conceptualization, formal analysis, supervision, validation, writing – original draft, writing – review & editing. All authors have read and approved the final version of the manuscript for publication.

Code and data availability

The datasets used in the numerical simulations throughout this article, as well as the code used to analyze them, can be found in the repository

`https://github.com/LaComarca-Lab/non-uniform-hypergraphs.`

Acknowledgments

Conflict of interest

All authors declare no conflicts of interest in this paper.

Prof. Miguel Romance is the Guest Editor of special issue "Complex Networks" for AIMS Mathematics. Prof. Miguel Romance was not involved in the editorial review and the decision to publish this article.

References

1. S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D. U. Hwang, Complex networks: Structure and dynamics, *Phys. Rep.*, **424** (2006), 175–308. https://doi.org/10.1016/j.physrep.2005.10.009

2. F. Battiston, G. Cencetti, I. Iacopini, V. Latora, M. Lucas, A. Patania, et al., Networks beyond pairwise interactions: Structure and dynamics, *Phys. Rep.*, **874** (2020), 1–92. https://doi.org/10.1016/j.physrep.2020.05.004

3. S. Boccaletti, P. De Lellis, C. del Genio, K. Alfaro-Bittner, R. Criado, S. Jalan, et al., The structure and dynamics of networks with higher-order interactions, *Phys. Rep.*, **1018** (2023), 1–64. https://doi.org/10.1016/j.physrep.2023.04.002

4. Z. K. Zhang, C. Liu, A hypergraph model of social tagging networks, *J. Stat. Mech.-Theory E.*, 2010. https://doi.org/10.1088/1742-5468/2010/10/P10005

5. X. L. Liu, C. Zhao, Eigenvector centrality in simplicial complexes of hypergraphs, *Chaos Interdisc. J. Nonlinear Sci.*, **33** (2023). https://doi.org/10.1063/5.0144871

6. L. Page, S. Brin, R. Motwani, T. Winograd, *The pagerank citation ranking: Bringing order to the web*, In: Proceedings of the 7th International World Wide Web Conference, 1998, 161–172.

7. A. R. Benson, Three hypergraph eigenvector centralities, *SIAM J. Math. Data Sci.*, **1** (2019), 293–312. https://doi.org/10.1137/18M1203031

8. L. Q. Qi, Z. Y. Luo, *Tensor analysis: Spectral theory and special tensors*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2017.

9. D. A. Bini, B. Meini, F. Poloni, On the solution of a quadratic vector equation arising in markovian binary trees, *Numer. Linear Algebr.*, **18** (2011), 981–991. https://doi.org/10.1002/nla.809

10. L. Q. Qi, Y. J. Wang, E. X. Wu, D-eigenvalues of diffusion kurtosis tensors, *J. Comput. Appl. Math.*, **221** (2008), 150–157. https://doi.org/10.1016/j.cam.2007.10.012

11. S. L. Hu, L. Q. Qi, G. F. Zhang, Computing the geometric measure of entanglement of multipartite pure states by means of non-negative tensors, *Phys. Rev. A*, **93** (2016). https://doi.org/10.1103/PhysRevA.93.012304

12. A. R. Benson, D. Gleich, J. Leskovec, *Tensor spectral clustering for partitioning higher-order network structures*, In: Proceedings of the 2015 SIAM International Conference on Data Mining, 2015, 118–126. https://epubs.siam.org/doi/abs/10.1137/1.9781611974010.14

13. M. Ng, L. Qi, G. L. Zhou, Finding the largest eigenvalue of a nonnegative tensor, *SIAM J. Matrix Anal. A.*, **31** (2009), 1090–1099. https://doi.org/10.1137/09074838X

14. S. Vigna, *Spectral ranking*, Cambridge University Press, 2016, 433–445. https://doi.org/10.1017/nws.2016.21

15. E. Estrada, *The structure of complex networks: Theory and applications*, Oxford University Press, New York, 2012.

16. C. D. Meyer, *Matrix analysis and applied linear algebra*, SIAM, 2023.

17. K. J. Pearson, T. Zhang, On spectral hypergraph theory of the adjacency tensor, *Graph. Combinator.*, **30** (2014), 1233–1248. https://doi.org/10.1007/s00373-013-1340-x

18. K. C. Chang, K. J. Pearson, T. Zhang, Some variational principles for z-eigenvalues of nonnegative tensors, *Linear Algebra Appl.*, **438** (2013), 4166–4182. https://doi.org/10.1016/j.laa.2013.02.013

19. A. R. Benson, D. Gleich, Computing tensor *z*-eigenvectors with dynamical systems, *SIAM J. Matrix Anal. A.*, **40** (2019), 1311–1324. https://doi.org/10.1137/18M1229584

20. S. G. Aksoy, I. Amburg, S. J. Young, Scalable tensor methods for nonuniform hypergraphs, *SIAM J. Math. Data Sci.*, **6** (2024). https://doi.org/10.1137/23M1584472

21. K. C. Chang, T. Zhang, On the uniqueness and non-uniqueness of the positive z-eigenvector for transition probability tensors, *J. Math. Anal. Appl.*, **408** (2013), 525–540. https://doi.org/10.1016/j.jmaa.2013.04.019

22. K. C. Chang, K. J. Pearson, T. Zhang, Perron-Frobenius theorem for nonnegative tensors, *Commun. Math. Sci.*, **6** (2008), 507–520.

23. K. Kovalenko, M. Romance, E. Vasilyeva, D. Aleja, R. Criado, D. Musatov, et al., Vector centrality in hypergraphs, *Chaos Soliton. Fract.*, **162** (2022), 112397. https://doi.org/10.1016/j.chaos.2022.112397

24. Y. M. Zhen, J. H. Wang, Community detection in general hypergraph via graph embedding, *J. Am. Stat. Assoc.*, **118** (2022), 1620–1629. https://doi.org/10.1080/01621459.2021.2002157

25. X. Ouvrard, J. M. Le Goff, S. Marchand-Maillet, Adjacency and tensor representation in general hypergraphs part 1: e-adjacency tensor uniformisation using homogeneous polynomials, *Arxiv preprint*, 2018. https://doi.org/10.48550/arXiv.1712.08189

26. A. Banerjee, A. Char, B. Mondal, Spectra of general hypergraphs, *Linear Algebra Appl.*, **518** (2017), 14–30. https://doi.org/10.1016/j.laa.2016.12.022

27. N. W. Landry, M. Lucas, I. Iacopini, G. Petri, A. Schwarze, A. Patania, et al., XGI: A Python package for higher-order interaction networks, *J. Open Source Softw.*, **8** (2023). https://doi.org/10.21105/joss.05162

28. A. R. Benson, R. Abebe, M. T. Schaub, A. Jadbabaie, J. Kleinberg, Simplicial closure and higher-order link prediction, *P. Natl. Acad. Sci.*, **115** (2018), E11221–E11230. https://doi.org/10.1073/pnas.1800683115

29. L. Isella, J. Stehlé, A. Barrat, C. Cattuto, J. F. Pinton, W. Van den Broeck, What's in a crowd? analysis of face-to-face behavioral networks, *J. Theor. Biol.*, **271** (2011), 166–180. https://doi.org/10.1016/j.jtbi.2010.11.033

30. D. R. Hofstadter, *Gödel, Escher, Bach: An Eternal Golden Braid*, Basic Books Inc., 1979.

31. J. Stehlé, N. Voirin, A. Barrat, C. Cattuto, L. Isella, J. F. Pinton, et al., High-resolution measurements of face-to-face contact patterns in a primary school, *PloS One*, **6** (2011), e23176. https://doi.org/10.1371/journal.pone.0023176

32. R. Mastrandrea, J. Fournet, A. Barrat, Contact patterns in a high school: A comparison between data collected using wearable sensors, contact diaries and friendship surveys, *PloS One*, **10** (2015), e0136497. https://doi.org/10.1371/journal.pone.0136497

33. C. Cattuto, W. Van den Broeck, A. Barrat, V. Colizza, J. F. Pinton, A. Vespignani, Dynamics of person-to-person interactions from distributed rfid sensor networks, *PloS One*, **5** (2010), e11596. https://doi.org/10.1371/journal.pone.0011596

34. K. I. Goh, M. E. Cusick, D. Valle, B. Childs, M. Vidal, A. L. Barabási, The human disease network, *P. Natl. Acad. Sci.*, **104** (2007), 8685–8690. https://doi.org/10.1073/pnas.0701361104

35. M. Dewar, J. Healy, X. Pérez-Giménez, P. Prałat, J. Proos, B. Reiniger, et al., Subhypergraphs in non-uniform random hypergraphs, *Arxiv preprint*, 2018. https://doi.org/10.48550/arXiv.1703.07686

36. T. G. Kolda, J. R. Mayo, An adaptive shifted power method for computing generalized tensor eigenpairs, *SIAM J. Matrix Anal. Appl.*, **35** (2014), 1563–1581. https://doi.org/10.1137/140951758

37. G. Gallo, G. Longo, S. Pallottino, S. Nguyen, Directed hypergraphs and applications, *Discrete Appl. Math.*, **42** (1993), 177–201. https://doi.org/10.1016/0166-218X(93)90045-P

38. G. Contreras-Aso, R. Criado, M. Romance, Beyond directed hypergraphs: Heterogeneous hypergraphs and spectral centralities, *J. Complex Netw.*, **12** (2024), cnae037. https://doi.org/10.1093/comnet/cnae037

39. J. L. Guo, X. Y. Zhu, Q. Suo, J. Forrest, Non-uniform evolving hypergraphs and weighted evolving hypergraphs, *Sci. Rep.*, **6** (2016), 36648. https://doi.org/10.1038/srep36648

AIMS Press