A framework for second-order eigenvector centralities and clustering coefficients

Citation for published version:

Digital Object Identifier (DOI):
10.1098/rspa.2019.0724

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and/or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
A framework for second-order eigenvector centralities and clustering coefficients

Francesca Arrigo¹, Desmond J. Higham ² and Francesco Tudisco³

¹University of Strathclyde, GX11H Glasgow (UK)
²University of Edinburgh, EH93FD Edinburgh (UK)
³Gran Sasso Science Institute, 67100 L’Aquila (Italy)

We propose and analyse a general tensor-based framework for incorporating second-order features into network measures. This approach allows us to combine traditional pairwise links with information that records whether triples of nodes are involved in wedges or triangles. Our treatment covers classical spectral methods and recently proposed cases from the literature, but we also identify many interesting extensions. In particular, we define a mutually-reinforcing (spectral) version of the classical clustering coefficient. The underlying object of study is a constrained nonlinear eigenvalue problem associated with a cubic tensor. Using recent results from nonlinear Perron–Frobenius theory, we establish existence and uniqueness under appropriate conditions, and show that the new spectral measures can be computed efficiently using a nonlinear power method. To illustrate the added value of the new formulation, we give an asymptotic analysis of a class of synthetic networks. We also give computational results on centrality and link prediction for real, large scale networks.

1. Introduction and motivation

The classical paradigm in network science is to analyse a complex system by focusing on pairwise interactions; that is, by studying lists of nodes and edges. However, it is now apparent that many important features arise through larger groups of nodes acting together [1]. For example, the triadic closure principle from the social sciences suggests that connected node triples, or triangles, are important building blocks [2–4]. Of course, there is a sense in which many algorithms in network science indirectly go beyond pairwise interactions by...
considering traversals around the network. However, recent work [4–7] has shown that there is benefit in directly taking account of higher-order neighbourhoods when designing algorithms and models.

From the point of view of algebraic topology, higher-order relations coincide with different homology classes and the idea of exploring connections of higher-order in networks is analogous to the idea of forming a filtered cell complex in topological data analysis [8]. Similarly to point clouds, complex networks modeling various type of interactions (such as social, biological, communication or food networks) have an intrinsic higher-order organization [1] and efficiently accounting for higher-order topology can allow more robust and effective quantification of the importance of nodes [9,10].

Our aim here is to develop and analyse a general framework for incorporating second-order features; see Definition 3.1. This takes the form of a constrained nonlinear eigenvalue problem associated with a cubic tensor. For specific parameter choices we recover both standard and recently proposed network measures as special cases, and we also construct many interesting new alternatives. In this eigenproblem-based setting, the network measures naturally incorporate mutual reinforcement; important objects are those that interact with many other important objects. The classic PageRank algorithm [11] is perhaps the best known example of such a measure. Within this setting, in Definition 3.2 we define for the first time a mutually reinforcing version of the classical Watts-Strogatz Clustering Coefficient [12]; here we give extra weight to nodes that form triangles with nodes that are themselves involved in important triangles. We show that our general framework can be studied using recent results from nonlinear Perron–Frobenius theory. As well as deriving existence and uniqueness results we show that these measures are computable via a nonlinear extension of the power method; see Theorem 4.1.

The manuscript is organized as follows. In section 2 we summarize relevant existing work on spectral measures in network science. Section 3 sets out the general framework for combining first and second order information through a tensor-based nonlinear eigenvalue problem. We also give several specific examples in order to show how standard measures can be generalized by including second order terms. In section 4 we study theoretical and practical issues. Section 5 illustrates the effect of using second order information through an asymptotic analysis on a specific class of networks. In section 6 we test the new framework on real large scale networks in the context of centrality assignment and link prediction. Conclusions are provided in section 7.

2. Background and related work
(a) Notation
A network or graph \( G = (V,E) \) is defined as a pair of sets: nodes \( V = \{1,2,\ldots,n\} \) and edges \( E \subseteq V \times V \) among them. We assume the graph to be undirected, so that for all \((i,j) \in E\) it also holds that \((j,i) \in E\), unweighted, so that all connections in the network have the same “strength”, and connected, so that it is possible to reach any node in the graph from any other node by following edges. We further assume for simplicity that the graph does not contain self-loops, i.e., edges that point from a node to itself.

A graph may be represented via its adjacency matrix, \( A = (A_{ij}) \in \mathbb{R}^{n \times n} \), where \( A_{ij} = 1 \) if \((i,j) \in E\) and \( A_{ij} = 0 \) otherwise. Under our assumptions, this matrix will be symmetric, binary and irreducible. We write \( G_A \) to denote the graph associated with the adjacency matrix \( A \).

We let \( 1 \in \mathbb{R}^n \) denote the vector with all components equal to 1 and \( 1_i \in \mathbb{R}^n \) denote the \( i \)th vector of the standard basis of \( \mathbb{R}^n \).

(b) Spectral centrality measures
A centrality measure quantifies the importance of each node by assigning to it a nonnegative value. This assignment must be invariant under graph isomorphism, meaning that relabelling
the nodes does not affect the values they are assigned. We focus here on degree centrality and a family of centrality measures that can be described via an eigenproblem involving the adjacency matrix. This latter family includes as special cases eigenvector centrality and PageRank.

The degree centrality of a node is found by simply counting the number of neighbours that it possesses; so node \( i \) is assigned the value \( d_i \), where \( d = A1 \). Degree centrality treats all connections equally, it does not take account of the importance of those neighbours. By contrast eigenvector centrality is based on a recursive relationship where node \( i \) is assigned a value \( x_i \geq 0 \) such that \( x \) is proportional to \( Ax \). We will describe this type of measure as mutually reinforcing, because it gives extra credit to nodes that have more important neighbours. Under our assumption that \( A \) is irreducible, the eigenvector centrality measure \( x \) corresponds to the Perron–Frobenius eigenvector of \( A \). We note that this measure was popularized in the 1970s by researchers in the social sciences, [13], but can be traced back to algorithms used in the 19th century for ranking chess players, [14]. For our purposes, it is useful to consider a general class of eigenvector based measures of the form

\[
x \geq 0 \quad \text{such that} \quad Mx = \lambda x,
\]

where \( M \in \mathbb{R}^{n \times n} \) is defined in terms of the adjacency matrix \( A \). For example, we may use the PageRank matrix

\[
M = cAD^{-1} + (1 - c)e1^T
\]

with \( c \in (0,1) \) and \( D \) the diagonal matrix such that \( D_{ii} = d_i \). With this choice, the eigenvector solution of (2.1) is the PageRank vector [11].

(c) Watts-Strogatz clustering coefficient

The Watts-Strogatz clustering coefficient was used in [15] to quantify an aspect of transitivity for each node. To define this coefficient, we use \( \triangle(i) = (A^4)_{ii}/2 \) to denote the number of unoriented triangles involving node \( i \). Note that node \( i \) is involved in exactly \( d_i(d_i - 1)/2 \) wedges centred at \( i \), that is, paths of the form \( hij \) where \( h, i, j \) are distinct. Hence node \( i \) can be involved in at most \( d_i(d_i - 1)/2 \) triangles. The local Watts–Strogatz clustering coefficient of node \( i \) is defined as the fraction of wedges that are closed into triangles:

\[
c_i = \begin{cases} 
\frac{2 \triangle(i)}{d_i(d_i - 1)} & \text{if } d_i \geq 2 \\
0 & \text{otherwise.}
\end{cases}
\]

It is easy to see that \( c_i \in [0,1] \) with \( c_i = 0 \) if node \( i \) does not participate in any triangle and \( c_i = 1 \) if node \( i \) has not left any wedges unclosed.

Related to this measure of transitivity for nodes there are two network-wide versions; the average clustering coefficient

\[
\bar{C} = \frac{1^Tc}{n} = \frac{1}{n} \sum_{i=1}^{n} c_i = \frac{2}{n} \sum_{i,d_i \geq 2} \frac{\triangle(i)}{d_i(d_i - 1)}
\]

and the global clustering coefficient or graph transitivity [16]:

\[
\hat{C} = \frac{6|K_3|}{\sum_i d_i(d_i - 1)},
\]

where \( |K_3| \) is the number of unoriented triangles in the network and the multiplicative factor of 6 comes from the fact that each triangle closes six wedges, i.e., the six ordered pairs of edges in the triangle. This latter measure has been observed to typically take values between 0.1 and 0.5 for real world networks; see [17]. The global and average clustering coefficients have been found to capture meaningful features and have found several applications [18,19];
I think two is OK.

however, they have been observed to behave rather differently for certain classes of networks [20]. In this work, on the other hand, we focus our attention on the local measure defined in (2.3), which describes a property of each individual node in the graph. Beyond social network analysis, this index has also found application, for example, in machine learning pipelines, where nodes features are employed to detect outliers [21] or to inform role discovery [22,23], in epidemiology, where efficient vaccination strategies are needed, and in and in psychology [24], where it is desirable to identify at-risk individuals.

We see from (2.3) that the Watts–Strogatz clustering coefficient may be viewed as a second-order equivalent of degree centrality in the sense that it is not mutually reinforcing—a node is not given any extra credit for forming triangles with well-clustered nodes. In Definition 3.2 below we show how a mutually reinforcing clustering coefficient can be defined.

3. General eigenvector model

To incorporate second order information, given a tensor $T \in \mathbb{R}^{n \times n \times n}$ and a parameter $p \in \mathbb{R}$ we define the operator $T_p : \mathbb{R}^n \to \mathbb{R}^n$ that maps the vector $x \in \mathbb{R}^n$ to the vector entrywise defined as:

$$T_p(x)_i = \sum_{j,k=1}^n T_{ijk} \mu_p(x_j, x_k),$$

(3.1)

where $\mu_p(a, b)$ is the power (or binomial) mean:

$$\mu_p(a, b) = \left( \frac{|a|^p + |b|^p}{2} \right)^{1/p}.$$

Recall that the following well known properties hold for $\mu$: i) $\lim_{p \to 0} \mu_p(a, b) = \mu_0$, where $\mu_0(a, b) = \sqrt{|ab|}$ is the geometric mean; ii) $\mu_{-1}(a, b) = 2(|x|^{-1} + |y|^{-1})^{-1}$ is the harmonic mean; iii) $\lim_{p \to +\infty} \mu_p = \max \{|a|, |b|\}$ is the maximum function; whereas $\lim_{p \to -\infty} \mu_p = \min \{|a|, |b|\}$ is the minimum.

We may then define the following nonlinear network operator, and associated spectral centrality measure, which combines first and second order interactions.

Definition 3.1. Let $\alpha \in \mathbb{R}$ be such that $0 \leq \alpha \leq 1$, let $p \in \mathbb{R}$ and let $M \in \mathbb{R}^{n \times n}$ and $T \in \mathbb{R}^{n \times n \times n}$ be an entrywise nonnegative square matrix and an entrywise nonnegative cubic tensor associated with the network, respectively. Define $M : \mathbb{R}^n \to \mathbb{R}^n$ as

$$M(x)_i = \alpha M x_i + (1-\alpha) T_p(x),$$

(3.2)

Then the corresponding first- and second-order eigenvector centrality of node $i$ is given by $x_i \geq 0$, where $x$ solves the constrained nonlinear eigenvalue problem

$$x \geq 0 \text{ such that } M(x) = \lambda x.$$  

(3.3)

If we set $\alpha = 1$ in (3.3) then only first-order interactions are considered, and we return to the classical eigenvector centrality measures discussed in section 2. Similarly, with $\alpha = 0$ only second-order interactions are relevant.

In the next subsection we discuss specific choices for $M$ and $T$.

We also note that in order for the measure in Definition 3.1 to be well defined, there must exist a unique solution to the problem (3.3). We consider this issue in section 4.
(a) Specifying \(M\) and \(T\)

In Definition 3.1, the matrix \(M\) should encode information about the first-order (edge) interactions, with the tensor \(T\) representing the triadic relationships among node triples, that is, second-order interactions.

Useful choices of \(M\) are therefore the adjacency matrix or the PageRank matrix (2.2). Another viable choice, which we will use in some of the numerical experiments, is a rescaled version of the adjacency matrix \(M = AD^{-1}\), which we will refer to as the random walk matrix.

We now consider some choices for the tensor \(T\) to represent second order interactions.

**Binary triangle tensor.** Perhaps the simplest choice of second order tensor is

\[
(T_B)_{ijk} = \begin{cases} 
1 & \text{if } i, j, k \text{ form a triangle} \\
0 & \text{otherwise}.
\end{cases} \tag{3.4}
\]

As discussed, for example, in [25], we can build \(T_B\) with worst case computational complexity of \(O(n^3)\) or \(O(m^{3/2})\). Moreover, in [26] the authors construct the triangles tensor of four large real-world networks (EMAIL EUALL, SOC EPINIONS1, WIKI TALK, TWITTER COMBINED) and observe that \(\text{nnz}(T) = O(6m)\).

Note also that this tensor is closely related to the matrix \(A \circ A^2\), where \(\circ\) denotes the componentwise product (also called the Hadamaard or Schur product), as shown in (3.10). It can be easily verified that, regardless of the choice of \(p\), \((T_B(1))_i = (A^3)_{ii} = 2\triangle(i)\) for all \(i \in V\).

**Random walk triangle tensor.** A “random walk” normalization of the tensor \(T\) in (3.4), which will be denoted by \(T_W \in \mathbb{R}^{n \times n \times n}\), is entrywise defined as

\[
(T_W)_{ijk} = \begin{cases} 
\frac{1}{\Delta(j,k)} & \text{if } i, j, k \text{ form a triangle} \\
0 & \text{otherwise},
\end{cases} \tag{3.5}
\]

where \(\Delta(j,k) = (A \circ A^2)_{jk}\) is the number of triangles involving the edge \((j, k)\). This is reminiscent of the random walk matrix \(M_{ij} = (AD^{-1})_{ij} = \delta_{ij}E/d_j\) (here \(\delta\) denotes the Kronecker delta) and this is the reason behind the choice of the name.

**Clustering coefficient triangle tensor.** A different normalization of the tensor in (3.4) is defined as

\[
(T_C)_{ijk} = \begin{cases} 
\frac{1}{\pi(d_i^{-1})} & \text{if } i, j, k \text{ form a triangle and } d_i \geq 2 \\
0 & \text{otherwise},
\end{cases} \tag{3.6}
\]

\[
\text{DJH} \diamond \text{I added the } d_i \geq 2 \text{ condition to this definition. Is this OK?}
\]

This tensor incorporates information that is not used in (3.4) and (3.5)—the number of transitive relationships that each node could be potentially involved in—while also accounting for the second-order structure actually present. We refer to (3.9) as the clustering coefficient triangle tensor because for any \(p\) we have \((T_C)_p(1) = c\), the WS clustering coefficient vector. We will return to this property in subsection (c).

**Local closure triangle tensor.** The local closure coefficient [27] of node \(i\) is defined as

\[
h_i = \frac{2\Delta(i)}{w(i)} \tag{3.7}
\]

where

\[
w(i) = \sum_{j \in N(i)} d_j - d_i = \sum_{j \in N(i)} (d_j - 1) \tag{3.8}
\]

is the number of paths of length two originating from node \(i\), and \(N(i)\) is the set of neighbours of node \(i\). We may also write \(w = A^2d - d = A^21 - A1\). The following result, which is an immediate consequence of the definition of \(w(i)\), shows that we may assume \(w(i) \neq 0\) when dealing with real-world networks.
Figure 1. Two toy network examples with the same number of edges (solid) and triangles. Left: node 1 can be involved in five more undirected triangles according to the principle underlying the WS clustering coefficient. These are formed using the dashed edges. Right: node 1 can only be involved in two more, formed using the dashed edges.

\[ (T_L)_{ijk} = \begin{cases} \frac{1}{\text{w}(i)} & \text{if } i, j, k \text{ form a triangle} \\ 0 & \text{otherwise} \end{cases} \] (3.9)

It is easily checked that \((T_L)_p(1) = h\) for all \(p\).

Next, we briefly discuss the main difference, for the purposes of this work, among the four tensorial network representations just introduced.

The binary triangle tensor (3.4) and random walk triangle tensor (3.5) provide no information concerning the wedges involving each node, and hence the consequent potential for triadic closure. Indeed, networks that have very different structures from the viewpoint of potential and actual transitive relationships are treated alike. For example, consider the two networks in Figure 1, where solid lines are used to represent the actual edges in the network. The two networks are represented by the same tensors in the case of (3.4) and (3.5), but are not alike from the viewpoint of transitive relationships. Indeed, by closing wedges in the network on the left-hand side node 1 could participate in five more triangles, whilst in the graph on the right-hand side it could participate in only two more. These are highlighted in Figure 1 using dashed lines.

On the other hand, the clustering coefficient triangle tensor defined in (3.6) encodes in its entries the “potential” for triadic closure of node 1; indeed, for the network on the left-hand side it holds that \((T_C)_{123} = (T_C)_{132} = 1/12\), while these entries are \((T_C)_{123} = (T_C)_{132} = 1/6\) for the network on the right-hand side. These values show that there is a potential for node 1 to be involved in respectively 12 and 6 directed triangles.

The local closure triangle tensor defined in (3.9) encodes another type of triadic closure property—the potential of a node to become involved in triangles by connecting to nodes that are at distance two from it. In the networks depicted in Figure 2 it is clear that no such triangles can be formed in the network in the left-hand side, while there is one that could be formed in the graph on the right-hand side (dashed edge). For the entries of the associated tensor \(T_L\), the left-hand network in Figure 2 has \((T_L)_{123} = (T_L)_{132} = 1/2\), and indeed node 1 is participating...
Figure 2. Two toy network examples with the same number of edges (solid) and triangles. Left: node 1 cannot be involved in any more triangles, according to the principle underlying the local closure coefficient. Right: node 1 can only be involved in one more, formed using the dashed edge.

in both possible directed triangles that can be formed according to the principals of local closure. The network on the right-hand side has $(T_L)_{123} = (T_L)_{132} = 1/3$.

FA ◊ The fact that $(T_L)_{123} = (T_L)_{132} = 1/3$ I initially found rather confusing. The explanation that I have (which may well be wrong) is as follows: It seems that the local closure coefficient accounts explicitly for the directionality of the triangle, thus it expects node 1 to be involved in $1 \rightarrow 2 \rightarrow 3 \rightarrow 1$, $1 \rightarrow 3 \rightarrow 2 \rightarrow 1$ and $1 \rightarrow 4 \rightarrow 5 \rightarrow 1$, but not in $1 \rightarrow 5 \rightarrow 4 \rightarrow 1$. This somehow makes sense because the measure closes paths of length two that originate at node 1. Therefore, the measure is “aware” that node 1 can reach node 5 in two steps, but it seems to be unable to then understand that adding $(1,5)$ would form two triangles rather than just one. It’s interesting because it is enforcing directionality in an undirected setting!

(b) The linear cases: $\alpha = 1$ or $p = 1$

The map $M$ defined in (3.2) becomes linear for particular choices of $p$ and $\alpha$. One case arises when $\alpha = 1$, whence it reduces to a standard matrix-vector product, $M(x) = Mx$, and (3.1) boils down to a linear eigenvector problem (2.1). Using the particular choices of $M$ described in the previous subsection, it then follows that our model includes as a special case standard eigenvector centrality and PageRank centrality.

Now let $\alpha \in [0, 1)$ and $p = 1$. Then the mapping $T_p : \mathbb{R}^n \rightarrow \mathbb{R}^n$ also becomes linear; indeed, entrywise it becomes

$$T_1(x)_i = \frac{1}{2} \sum_{j,k=1}^n T_{ijk} x_k + T_{ijk} x_j = \frac{1}{2} \left\{ \sum_{j=1}^n (\sum_{k=1}^n T_{ikj}) x_j + \sum_{j=1}^n (\sum_{k=1}^n T_{ijk}) x_j \right\}$$

and $T_1(x)$ reduces to the product between the vector $x$ and the matrix with entries $\frac{1}{2}(\sum_k T_{ijk} + T_{ikj})$. In particular, if the tensor $T$ is symmetric with respect to the second and third modes, i.e. $T_{ijk} = T_{ikj}$ for all $j, k$, it follows that

$$T_1(x)_i = \sum_{j=1}^n (\sum_{k=1}^n T_{ijk}) x_j.$$

Note that this is the case for all the tensors defined in subsection (a).

We now explicitly compute $(\sum_k T_{ijk})$ for some of the tensors $T$ presented in subsection (a). If $T = T_B$ is the binary triangle tensor in (3.4), it follows that

$$\sum_{k=1}^n (T_B)_{ijk} = (A \circ A^2)_{ij}$$

and hence

$$(T_B)_1(x) = (A \circ A^2)x.$$
Overall, the map $\mathcal{M}$ then acts on a vector $\mathbf{x}$ as follows

$$
\mathcal{M}(\mathbf{x}) = \alpha \mathbf{M} \mathbf{x} + (\mathbf{T}_B)_1(\mathbf{x}) = \left(\alpha A + (1 - \alpha)(A \circ A^2)\right) \mathbf{x},
$$

and so the solution to the constrained eigenvector problem (3.3) is the Perron–Frobenius eigenvector of the matrix $\alpha A + (1 - \alpha)(A \circ A^2)$. This has a flavour of the work in [1], where the use of $A \circ A^2$ is advocated as a means to incorporate motif counts involving second-order structure. Other choices of the tensor $\mathbf{T}$ yield different eigenproblems. For example, when $\mathbf{T} = \mathbf{T}_C$ in (3.6) we have

$$
\sum_{k=1}^{n} (\mathbf{T}_C)_{ijk} = \begin{cases} (\alpha \circ A^2)_{ij} \overline{d_i} & \text{if } d_i \geq 2 \\ 0 & \text{otherwise} \end{cases}
$$

and hence (3.2) becomes

$$
\mathcal{M}(\mathbf{x}) = \alpha \mathbf{M} \mathbf{x} + (1 - \alpha)(\mathbf{T}_C)_1(\mathbf{x}) = \left(\alpha A + (1 - \alpha)(D^2 - D)\right) (\alpha \circ A^2) \mathbf{x},
$$

where $\dagger$ denotes the Moore-Penrose pseudo-inverse. If we let $\mathbf{T} = \mathbf{T}_L$, as defined in (3.9), we obtain

$$
\sum_{k=1}^{n} (\mathbf{T}_L)_{ijk} = \begin{cases} (\alpha \circ A^2)_{ij} \overline{w(i)} & \text{if } d_j \geq 2 \\ 0 & \text{otherwise} \end{cases}
$$

(3.11)

Note that, in formula (3.11), $w(i) = 0$ may hold for some $i$ even though $d_j \geq 2$. However, as observed in Proposition 3.1, in that case $i$ cannot form any triangle and thus $(\alpha \circ A^2)_{ij} = 0$ and we have $\sum_{k=1}^{n} (\mathbf{T}_L)_{ijk} = 0$ as well. Using (3.11) we obtain

$$
(\mathbf{T}_L)_1(\mathbf{x}) = W^\dagger (\alpha \circ A^2) \mathbf{x},
$$

where $W = \text{diag}(w(1), \ldots, w(n))$. The eigenvector problem (3.3) then becomes

$$
\mathcal{M}_{1,\alpha}(\mathbf{x}) = \alpha \mathbf{M} \mathbf{x} + (1 - \alpha)(\mathbf{T}_L)_1(\mathbf{x}) = \left(\alpha A + (1 - \alpha)W^\dagger (\alpha \circ A^2)\right) \mathbf{x} = \lambda \mathbf{x}.
$$

(c) Spectral clustering coefficient: $\alpha = 0$

While the choice of $\alpha = 1$ yields a linear and purely first-order map, the case $\alpha = 0$ corresponds to a map that only accounts for second-order node relations. In particular, this map allows us to define spectral, and hence mutually reinforcing, versions of the Watts–Strogatz clustering coefficient (2.3) and the local closure coefficient (3.8). We therefore make the following definition.

**Definition 3.2.** Let $\mathbf{T} \in \mathbb{R}^{n \times n \times n}$ be an entrywise nonnegative cubic tensor associated with the network. The spectral clustering coefficient of node $i$ is the $i$th entry of the vector $\mathbf{x} \geq 0$ which solves the eigenvalue problem (3.3) with $\alpha = 0$ in (3.2); that is,

$$
T_p(x)_i = \lambda x_i.
$$

(3.12)

The solution for $\mathbf{T} = \mathbf{T}_C \in \mathbb{R}^{n \times n \times n}$ in (3.6) will be referred to as the spectral Watts–Strogatz clustering coefficient, and the solution for $\mathbf{T} = \mathbf{T}_L \in \mathbb{R}^{n \times n \times n}$ in (3.9) will be referred to as the spectral local closure coefficient.

The power mean parameter $p$ in (3.1) controls how the clustering coefficients of neighbouring nodes are combined. However, some properties are shared by all possible choices of $p$. In particular, for all values of $p$ we have $(\mathbf{T}_C)_0(1) = c$ and $(\mathbf{T}_L)_0(1) = h$, where $c$ and $h$ are the vectors of Watts–Strogatz clustering coefficients and local closure coefficients, as defined in (2.3) and (3.8), respectively.

More generally, $T_p(1)$ defines a “static” counterpart of the spectral clustering coefficient obtained as the Perron–Frobenius eigenvector of $\mathbf{T}_p$. This may be viewed as a second-order analogue of the dichotomy between degree vector and eigenvector centrality, the former being
defined as $A1$ and the latter as the Perron–Frobenius eigenvector of $A$. As in the first-order case, even though the spectral coefficient $x \propto T_p(x)$ carries global information on the network while the static version $T_p(1)$ is highly local, the two measures can be correlated. An example of this phenomenon is shown in Figure 3, which scatter plots $T_p(1)$ against $T_0(x)$, for different choices of $T$, on the unweighted version of the neural network of C. Elegans compiled by Watts and Strogatz in [15], from original experimental data by White et al. [28]. In Table 1 we summarize the number of nodes $n$, edges $m$, triangles $\triangle$ and the global clustering coefficient $\hat{C}$ of this network.

We also remark that our general definition of spectral clustering coefficient in Definition 3.2 includes in the special case $p \to 0$ the Perron $H$-eigenvector of the tensor $T$ [29]. Indeed, it is easy to observe that the change of variable $y^2 = x$ yields

$$T_0(x) = \lambda x \iff T_{yy} = \lambda y^2,$$

where $T_{yy}$ is the tensor-vector product $(T_{yy})_i = \sum_j T_{ijk} y_j y_k$. This type of eigenvector has been used in the context of hypergraph centrality in, for example, [6].

Note that if node $i$ is not part of any triangle, then the summation describing the corresponding entry in $T_p(x)$ is empty, and thus the spectral clustering coefficient for this node is zero, as expected. Moreover, the converse is also true, since $T \geq 0$ and $x \geq 0$. On the other hand, since the spectral clustering coefficient $x$ is defined via an eigenvector equation for $T_p$, it follows that it cannot be unique as it is defined only up to a positive scalar multiple. Indeed, we have $T_p(\beta x) = \beta T_p(x)$ for any $\beta \geq 0$. Hence, when $T = T_C$, unlike the standard Watts-Strogatz clustering coefficient, it is no longer true that a unit spectral clustering coefficient identifies nodes that participate in all possible triangles. However, we will see in the next section that once we have a solution $x$ of (3.12), then any other solution must be a positive multiple of $x$. More precisely, we will show that under standard connectivity assumptions on the network, the spectral clustering coefficient and, more generally, the solution to (3.3) is unique up to positive scalar multiples. This fosters the analogy with the linear setting (2.1). Therefore, it is meaningful to normalize the solution to (3.3) and compare the size of its components to infer information on the relative importance of nodes within the graph.

\footnote{We are binarizing the original weighted network by assigning weight one to every edge.}
The choice of the tensor $T$ affects the way the triangle structure is incorporated in our measure. An example of the differences that one may obtain is displayed in Figure 4. Here we plot the Karate network and highlight the ten nodes which score highest according to the spectral clustering coefficient for different choices of $T$. This is a social network representing friendships between the 34 members of a karate club at a US university [30]; see Table 1 for the network statistics. In this experiment we select $p = 0$, and thus we are actually computing the Perron $H$-eigenvector of the corresponding tensors. The size of each of the top ten nodes in Figure 4 is proportional to their clustering coefficients. In Figure 5, instead, we show how the $H$-eigenvectors corresponding to different triangle tensors correlate with the degree of the nodes. We group nodes by logarithmic binning their degree and plot the average degree versus the average clustering coefficients in each bin. As expected, the Watts–Strogatz spectral clustering coefficient may decrease when the degree increases, in contrast with other choices of the triangle tensor. A similar phenomenon is observed for example in [27].

In the next section we proceed to discuss existence and uniqueness, up to scalar multiples, of a solution to (3.3). We also describe a power-iteration algorithm for its computation.
Theorem 4.1. If moreover $M$ is aperiodic, then there exists a unique (up to multiples) nonnegative eigenvector of $M$, and such vector is entrywise positive. Moreover, this eigenvector is maximal, since the corresponding eigenvalue is the spectral radius of $B$ and, if $G_B$ is aperiodic, the power method iteration $x_{k+1} = Bx_k / \|Bx_k\|$ converges to it for any starting vector $x_0 \in \mathbb{R}^n$.

In the general case, we will appeal to nonlinear Perron–Frobenius theory to show that the properties of existence, uniqueness and maximality of the solution to (3.3) carry over to the general nonlinear setting almost unchanged, and to show that an efficient iteration can be used to compute this solution. We first note that for any $\alpha \in [0, 1]$, any $p \in \mathbb{R}$ and any $\theta > 0$ we have

$$M(\theta x) = \alpha M(\theta x) + (1 - \alpha) T_p(\theta x) = \theta M(x),$$

thus if $x \geq 0$ solves (3.3), then any positive multiple of $x$ does as well. Therefore, as for the linear case, uniqueness can only be defined up to scalar multiples. We continue by introducing the graph of $M$.

Definition 4.1. Given a matrix $M \in \mathbb{R}^{n \times n}$ and a cubic tensor $T \in \mathbb{R}^{n \times n \times n}$, both assumed to be nonnegative, we define the adjacency graph $G_M$ of $\mathcal{M}$ in (3.2) as the pair $G_M = (V, E_M)$ where $V = \{1, \ldots, n\}$ and, for all $i, j \in V$, $(i, j) \in E_M$ if and only if $(A_M)_{ij} = 1$, where $A_M$ is the adjacency matrix entrywise defined as

$$(A_M)_{ij} = \begin{cases} 1 & \text{if } \alpha M_{ij} + (1 - \alpha) \sum_{k=1}^n (T_{ijk} + T_{ikj}) > 0 \\ 0 & \text{otherwise} \end{cases} \quad (4.1) \tag{eq:graphM}$$

We now state and prove our main theorem.

Theorem 4.1. Given the nonnegative matrix $M \in \mathbb{R}^{n \times n}$ and the nonnegative tensor $T \in \mathbb{R}^{n \times n \times n}$, let $\mathcal{M}$ be defined as in (3.2) and let $G_M$ be its adjacency graph, as in Definition 4.1. If $G_M$ is strongly connected, then

(i) There exists a unique (up to multiples) positive eigenvector of $\mathcal{M}$, i.e. a unique positive solution of (3.3).

(ii) The positive eigenvector of $\mathcal{M}$ is maximal, i.e. its eigenvalue is $\rho(\mathcal{M}) = \max\{|\lambda| : \mathcal{M}(x) = \lambda x\}$.

(iii) If $x$ is any nonnegative eigenvector $\mathcal{M}(x) = \lambda x$ with some zero entry, then $\lambda < \rho(\mathcal{M})$.

If moreover $G_M$ is aperiodic, then

(iv) For any starting point $x_0 \geq 0$, the nonlinear power method

$$\begin{align*}
y_{k+1} &= \alpha Mx_k + (1 - \alpha) T_p(x_k) \\
x_{k+1} &= y_{k+1} / \|y_{k+1}\|
\end{align*}$$

converges to the positive eigenvector of $\mathcal{M}$. Moreover, for all $k = 0, 1, 2, \ldots$ it holds

$$\min_{i=1, \ldots, n} (y_{k+1})_i \leq \min_{i=1, \ldots, n} (y_{k})_i \leq \rho(\mathcal{M}) \leq \max_{i=1, \ldots, n} (y_{k+1})_i \leq \max_{i=1, \ldots, n} (y_{k})_i \quad (4.2) \tag{eq:conv}$$

with both the left and the right hand side sequences converging to $\rho(\mathcal{M})$ as $k \to \infty$.

Proof. The proof combines several prior results from nonlinear Perron–Frobenius theory.
First, note that $\mathcal{M}$ is homogeneous of degree one and order preserving. Indeed, if $x \geq y \geq 0$ entrywise, then it is easy to verify that

$$\mathcal{M}(x) = \alpha \mathcal{M} x + (1 - \alpha) T_p(x) \geq \alpha \mathcal{M} y + (1 - \alpha) T_p(y) = \mathcal{M}(y) \geq 0.$$  

Thus $\mathcal{M}$ has at least one entrywise nonnegative eigenvector that corresponds to the eigenvalue $\lambda = \rho(\mathcal{M})$ (see, e.g., [31, Theorem 5.4.1]).

Second, we may use [32, Theorem 1] to show that there exists at least one positive eigenvector of $\mathcal{M}$. Let $1_j$ denote the $j$th vector of the canonical basis of $\mathbb{R}^n$. Then let $y_j(\beta) = 1 + (\beta - 1) 1_j$ be the vector whose $j$th component is the variable $\beta \in \mathbb{R}$ while all the other entries are equal to one. Thus note that if $A_{ij} = 1$, then $\lim_{\beta \to \infty} \mathcal{M}(y_j(\beta)) = \infty$. Since $G_{\mathcal{M}}$ is strongly connected, [32, Theorem 1] implies that $\mathcal{M}$ has at least one entrywise positive eigenvector $u > 0$ such that $\mathcal{M}(u) = \lambda u$, with $\lambda > 0$.

Third, we show uniqueness and maximality. Note that for any positive vector $y > 0$ and any $p \geq 0$ we have that if $G_{\mathcal{M}}$ is strongly connected then the Jacobian matrix of $\mathcal{M}$ evaluated at $y$ is irreducible. In fact

$$\frac{\partial}{\partial y_j} \mathcal{M}(y)_i = \alpha A_{ij} + (1 - \alpha) y_j^{-p-1} \sum_k (T_{ijk} + T_{ikj}) \mu_j(y_j, y_k)^{1-p}.$$  

Therefore, [31, Theorem 6.4.6] implies that $u$ is the unique positive eigenvector of $\mathcal{M}$. Moreover, [31, Theorem 6.1.7] implies that for any nonnegative eigenvector $x \geq 0$ with $\mathcal{M}(x) = \lambda x$ we have $\lambda \leq \rho(\mathcal{M})$. As there exists at least one nonnegative eigenvector corresponding to the spectral radius, that must be $u$ and we deduce that $\lambda = \rho(\mathcal{M})$.

This proves points (i) – (iii). For point (iv), we note that if $G_{\mathcal{M}}$ is aperiodic then $A_{\mathcal{M}}$ is primitive and this implies that the Jacobian matrix of $\mathcal{M}$ evaluated at $u > 0$ is primitive as well. Thus Theorem 6.5.6 and Lemma 6.5.7 of [31] imply that the normalized iterates of the homogeneous and order preserving map $\mathcal{M}$ converge to $u$. Finally, [33, Theorem 7.1] proves the sequence of inequalities in (4.2) and the convergence of both the sequences

$$\alpha_k = \min_{i=1,\ldots,n} \frac{M(x_k)_i}{(x_k)_i} \quad \text{and} \quad \beta_k = \max_{i=1,\ldots,n} \frac{M(x_k)_i}{(x_k)_i}$$

towards the same limit $\lim_{k} \alpha_k = \lim_{k} \beta_k = \rho(\mathcal{M})$.

To confirm the relevance of this result, the next lemma shows that for all choices of $M$ and $T$ in subsection (a), the graph $G_{\mathcal{M}}$ is undirected and coincides with the underlying network. Thus we may conclude that Theorem 4.1 applies whenever the original graph is connected.

**Lemma 4.1.** Let $\alpha \neq 0$ and $\mathcal{M}$ and $T$ be defined according to any of the choices in subsection 3(a). Then $\mathcal{M}$ and $A_{\mathcal{M}}$ have the same sparsity pattern; that is, $M_{ij} > 0$ if and only if $(A_{\mathcal{M}})_{ij} = 1$.

**Proof.** If $(i, j) \in E$ is an edge in the graph associated with $M$, i.e., $M_{ij} > 0$, then clearly $(A_{\mathcal{M}})_{ij} = 1$ as the tensor $T$ has nonnegative entries. If $(i, j) \notin E$, then from the possible definitions of the tensor $T$ listed in Subsection 3(a) it follows that $T_{ijk} = T_{ikj} = 0$, for all $k$. Thus $(A_{\mathcal{M}})_{ij} = M_{ij} = 0$. Vice versa, if $(A_{\mathcal{M}})_{ij} = 1$, then $\alpha M_{ij} + (1 - \alpha) \sum_{k=1}^{n} (T_{ijk} + T_{ikj}) = 0$. Since we are summing two nonnegative terms, it follows that both are zero and, in particular, $M_{ij} = 0$. If $(A_{\mathcal{M}})_{ij} = 1$, on the other hand, this implies $\alpha M_{ij} + (1 - \alpha) \sum_{k=1}^{n} (T_{ijk} + T_{ikj}) > 0$ and hence at least one of the two terms has to be positive; however, from the possible definitions of $T$ it is clear that $T_{ijk}$ and $T_{ikj}$ cannot be nonzero unless $(i, j) \in E$, i.e., unless $M_{ij} > 0$.

**5. Example network with theoretical comparison**

In this section we describe theoretical results on the higher order centrality measures. Our overall aim is to confirm that the incorporation of second order information can make a qualitative difference to the rankings. We work with networks of the form represented in Figure 6. These have three different types of nodes: i) node 1, the center of the wheel, that has degree $m$ and
connects to \( m \) nodes of the second type, ii) \( m \) nodes attached to node 1 and interconnected via a cycle to each other. Each type (ii) node also connects to \( k \) nodes of the third type, and iii) \( mk \) leaf nodes attached in sets of \( k \) to the \( m \) nodes of type (ii). Figure 6 shows the case where \( m = 5 \) and \( k = 3 \). We will use node 2 to represent the nodes of type (ii) and node \( m + 2 \) to represent the nodes of type (iii).

The network is designed so that node 1 is connected to important nodes and is also involved in many triangles. Node 2, by contrast, is only involved in two triangles and has connections to the less important leaf nodes. If we keep \( m \) fixed and increase the number of leaf nodes, \( k \), connected to node 2, then eventually we would expect the centrality of node 2 to overtake that of node 1. We will show that this changeover happens for a larger value of \( k \) when we incorporate second order information. Our goal is to show that including the higher order information drawn from the triangle. More precisely, we set \( p = 1 \) and show that node 1 is identified by the higher-order measure as being more central than node 2 for larger values of \( k \) when compared with standard eigenvector centrality.

With this labeling of the nodes, the adjacency matrix \( A \in \mathbb{R}^{M \times M} \) of the network has the form

\[
A = \begin{bmatrix}
0 & 1^T_m & 0 & \cdots & 0 \\
1_m & C & I_m \otimes 1^T_k \\
0 & \vdots & I_m \otimes 1^T_k \\
0 & 0 & 0 & \cdots & 0
\end{bmatrix}, \quad C = \begin{bmatrix}
0 & 1 & 1 \\
1 & \ddots & \ddots \\
1 & \ddots & 1 \\
1 & 1 & 0
\end{bmatrix} \in \mathbb{R}^{m \times m}.
\]

The unit 2-norm eigenvector associated to the leading eigenvalue \( \lambda = 1 + \sqrt{1 + m + k} \) of \( A \) has the form \( v = [x \, y \, 1^T_m \, z^T_{mk}]^T \), where

\[
x = \frac{m}{\sqrt{m(m + \lambda^2 + k)}} \quad y = \frac{\lambda}{\sqrt{m(m + \lambda^2 + k)}} \quad z = \frac{1}{\sqrt{m(m + \lambda^2 + k)}},
\]

and thus \( x > y \) if and only if \( m > \lambda \). It can be verified that this is equivalent to requiring

\( k < m(m - 3) \).
We now move on to the higher order setting. We begin by specifying the tensor $T = (T_{ijk})$ defined in (3.4). It is clear that $T_{ijk} = 0$ for all $i = m + 2, \ldots, mk + m + 1$. Moreover,

$$T_{1jk} = \begin{cases} 1 & \text{if } j, k = 2, \ldots, m + 1 \text{ are such that } (j, k) \in E \\ 0 & \text{otherwise}, \end{cases}$$

and for $i = 2, \ldots, m + 1$

$$T_{ijk} = \begin{cases} 1 & \text{if either } j = 1 \text{ and } (i, k) \in E \text{ or } k = 1 \text{ and } (i, j) \in E \\ 0 & \text{otherwise} \end{cases}$$

Using the definition of $T_p(v)$ it follows that, if we call $v = [x, y, 1_m^T, z1_{mk}^T]^T$ then

$$(T_p(v))_1 = 2my, \quad (T_p(v))_2 = 4\mu_p(x, y), \quad (T_p(v))_{m+2} = 0$$

and hence, for $\gamma = 1$, equation (3.3) boils down to

$$\begin{cases} \lambda x = (2 - \alpha)my \\ \lambda y = (\alpha + 2y + k) + 4(1 - \alpha)\mu_p(x, y) \\ \lambda z = \alpha y \\ + \text{normalization}. \end{cases}$$

Imposing $p = 1$, after some algebraic manipulation it turns out that $x > y$ for $\alpha \neq 0$ if and only if

$$k < \frac{(2 - \alpha)}{\alpha^2} \left( (2 - \alpha)m^2 + (\alpha - 4)m \right).$$

The areas for which $x > y$ in the two settings (standard eigenvector centrality $\alpha = 1$ and higher order centrality $\alpha = 0.2, 0.5$) are shaded in Figure 7 (left). It is readily seen that even for small values of $m, k$ needs to become very large (when compared to $m$) in order for the centrality of nodes $i = 2, \ldots, m + 1$ to become larger than that of node 1 when higher-order information is taken into account. In the standard eigenvector centrality setting we observe a very different behaviour (see Figure 7, left, $\alpha = 1$).

In Figure 7 (right) we display the areas for which $x > y$ for different values of $\alpha$ when $T_C \in \mathbb{R}^{n \times m \times k}$ is used in (3.3). Indeed, specializing the definition in (3.6) to this example, it is easy to see that

$$(T_C)_1 = \frac{2y}{m-1}, \quad (T_C)_2 = \frac{4\mu_p(x, y)}{(k + 3)(k + 2)}, \quad (T_C(m+2) = 0,$$

and the solution to (3.3) has to satisfy

$$\begin{cases} \lambda x = (\alpha m + \frac{2(1 - \alpha)}{m-1})y \\ \lambda y = \alpha (x + 2y + k\epsilon) + \frac{4(1 - \alpha)}{k + 3}(k + 2)\mu_p(x, y) \\ \lambda z = \alpha y \\ + \text{normalization}. \end{cases}$$

Then $x > y$ if and only if $\alpha m + \frac{2(1 - \alpha)}{m-1} > \lambda$, where, for $p = 1$, $\lambda$ satisfies

$$\lambda^2 - (2\alpha + c_1)\lambda - (\alpha + c_1)(\alpha m + c_2) - k\alpha^2 = 0$$

with $c_1 = \frac{2(1 - \alpha)}{(k + 3)(k + 2)}$ and $c_2 = \frac{2(1 - \alpha)}{m-1}$.

Remark 5.1. Working with $T_L$ leads to $x > y$ if and only if $\alpha m + \frac{2(1 - \alpha)}{m + 2k + 3} > \lambda$ where now $\lambda$ satisfies (5.1) for $c_1 = \frac{2(1 - \alpha)}{m + 2k + 3}$ and $c_2 = \frac{2(1 - \alpha)}{k + 2}$. There seems to be no appreciable difference between the profiles for $\alpha = 0.2, 0.5, 1$. 
6. Applications and numerical results

(a) Centrality measures

In the previous subsection we observed that different values of $\alpha$ led to different node rankings. Results were shown for $T_B$ and $T_C$, $p = 1$ and $\alpha = 0.1, 0.5, 1$. In this subsection we test those findings on real network data. We use $\alpha = 0.5$ and $\alpha = 1$ (corresponding to eigenvector centrality) and $p = 0$ in (3.3), and combine the adjacency matrix $A$ and the binary tensor $T_B$. The centrality vectors have been normalized in the infinity norm. The tests were performed on four real-world networks, available online at [34]. In Table 1 we report the number of nodes $n$, (undirected) edges $m$ and of triangles $\Delta = \text{trace}(A^3)/6$, and the global clustering coefficient $\hat{C}$ for the four networks. We further display the average clustering coefficient $\bar{c}$ and the average spectral clustering coefficient $\bar{x}_C$ as well as the average local closure coefficient $\bar{w}$ [27] and its spectral counterpart $\bar{x}_L$; see Definition 3.2. The two spectral measures were both computed with $p = 0$ and the solution to (3.12) was normalized in the infinity norm.
Table 1. Description of the dataset: \( n \) is the number of nodes, \( m \) is the number of edges, \( \Delta \) is the number of triangles, \( \hat{C} \) is the global clustering coefficient of the network, as defined in (2.4), \( \bar{\tau} \) is the average clustering coefficient, \( \bar{\pi}_C \) is the average spectral clustering coefficient, \( \bar{\pi}_w \) is the average local closure coefficient, and \( \bar{\pi}_L \) is the average spectral local closure coefficient.

<table>
<thead>
<tr>
<th>Name</th>
<th>( n )</th>
<th>( m )</th>
<th>( \Delta )</th>
<th>( \hat{C} )</th>
<th>( \bar{\tau} )</th>
<th>( \bar{\pi}_C )</th>
<th>( \bar{\pi}_w )</th>
<th>( \bar{\pi}_L )</th>
</tr>
</thead>
<tbody>
<tr>
<td>KARATE</td>
<td>34</td>
<td>78</td>
<td>45</td>
<td>0.26</td>
<td>0.57</td>
<td>0.12</td>
<td>0.22</td>
<td>0.23</td>
</tr>
<tr>
<td>CHESAPEAKE</td>
<td>39</td>
<td>170</td>
<td>194</td>
<td>0.28</td>
<td>0.45</td>
<td>0.41</td>
<td>0.25</td>
<td>0.38</td>
</tr>
<tr>
<td>ADJNOUN</td>
<td>112</td>
<td>425</td>
<td>284</td>
<td>0.16</td>
<td>0.17</td>
<td>0.18</td>
<td>0.09</td>
<td>0.18</td>
</tr>
<tr>
<td>C. ELEGANS</td>
<td>277</td>
<td>1918</td>
<td>2699</td>
<td>0.19</td>
<td>0.28</td>
<td>0.05</td>
<td>0.15</td>
<td>0.20</td>
</tr>
</tbody>
</table>

All networks in the dataset are undirected and unweighted, so C. ELEGANS has been symmetrized and made binary.

The selected real-world networks are often used as benchmarks in the graph clustering and community detection communities [34]. The KARATE network is a social network while C. ELEGANS is a neural network, as described in section (c). The network ADJNOUN is based on common adjective and noun adjacencies in the novel “David Copperfield” by Charles Dickens [35]. CHESAPEAKE represents the interaction network of the Chesapeake Bay ecosystem. Here, nodes represent species or suitably defined functional groups and links create the food web [36].

Figure 8 scatter plots the newly introduced measure against eigenvector centrality for the four different networks. As the figure shows, for the network KARATE there is very poor correlation between the two measures. Stronger correlation is displayed for the other networks, but it is still to be noted that the top ranked nodes (corresponding to the nodes with largest centrality scores) differ for the two measures in all but one network, namely ADJNOUN. Hence, using second-order information can alter our conclusions about which nodes are the most central.

(b) Link Prediction

Link prediction is a fundamental task in network analysis: in this setting we are given a network \( G_0 = (V, E_0) \) and asked to identify edges, i.e., pairs of nodes, that are not in \( E_0 \) but should be there. This problem typically arises in two settings: (a) in a dynamic network where new connections appear over time, and (b) in a noisily observed network, where it is suspected that edges are missing [37–39].

For convenience, let us assume that \( E_0 \) is the set of edges that we observe and that \( E_1 \) with \( E_1 \cap E_0 = \emptyset \) is the set of edges that should be predicted, i.e., those that will appear in an evolving network or that are missing in a noisy graph. A standard approach for link prediction is to create a similarity matrix \( S \), whose entries \( S_{ij} \) quantify the probability that \( (i, j) \in E_1 \). It is worth pointing...
out that since \( E_0 \cap E_1 = \emptyset \), then the nonzero pattern of \( S \) will be complementary to that of the adjacency matrix of \( G_0 \). Over the years, several similarity measures have been proposed in order to quantify which nodes are most likely to link to a given node \( i \). While classical methods usually exploit the first-order structure of connections around \( i \), there is a growing interest in second-order methods that take into account, for example, triangles.

In this context, we propose a new similarity measure based on \( M \) and its Perron eigenvector. This measure is a generalization of the well-known technique known as seeded (or rooted) PageRank \([40,41]\), which we now describe. Given a seed node \( \ell \in V \) and the teleportation coefficient \( 0 \leq c < 1 \), let \( x^{(\ell)} \) be the limit of the evolutionary process

\[
x_{k+1} = cP x_k + (1 - c)1_\ell, \quad k = 0, 1, 2, \ldots
\]

(6.1) where \( P \) is the random walk matrix \( P = AD^{-1} \). As \( 0 \leq c < 1 \), it is easy to show that the limit exists and that it coincides with the solution to the linear system

\[
(I - cP)x^{(\ell)} = (1 - c)1_\ell.
\]

(6.2) The seeded PageRank similarity matrix \( S_{PR} \) is then entrywise defined by

\[
(S_{PR})_{ij} = (\mathbf{x}_{(\ell)})_i + (\mathbf{x}_{(j)})_i.
\]

The idea behind (6.1) is that the sequence \( x_k \) is capturing the way a unit mass centered in \( \ell \) (the seed or root of the process), and represented in the model by \( 1_\ell \), propagates through the network following the diffusion rule described by \( P \). This diffusion map is a first-order random walk on the graph.

In order to propose a new, second-order, similarity measure, we replace this first-order map with the second-order diffusion described by \( M = \alpha M + (1 - \alpha)T_P \) and we consider the associated diffusion process. We begin by observing that independently of the choice of the starting point \( x_0 \) in (6.1), this diffusion process will always converge to \( x^{(\ell)} \) that satisfies \( \|x^{(\ell)}\|_1 = 1 \). Indeed, (6.2) yields

\[
\|x^{(\ell)}\|_1 = (1 - c) \sum_{k \geq 0} c^k P^k 1_\ell = (1 - c) \sum_{k \geq 0} c^k = 1.
\]

As a consequence, the limit of the sequence (6.1) coincides with the limit of the normalized iterates \( \hat{x}_{k+1} = cP \hat{x}_k + (1 - c)1_\ell \), with \( \hat{x}_{k+1} = \hat{x}_{k+1}/\|\hat{x}_{k+1}\|_1 \). On the other hand, when the linear process \( P \) is replaced by the nonlinear map \( M \), the unnormalized sequence may not converge. We thus need to impose normalization of the vectors in our dynamical process defined in terms of \( M \) and seeded in node \( \ell \):

\[
y_{k+1} = cM(y_k) + (1 - c)1_\ell \quad k = 0, 1, 2, \ldots
\]

(6.3) Note that, for \( \alpha = 1 \) and \( M = P \) in (3.2) we retrieve exactly the rooted PageRank diffusion (6.1). Unlike the linear case, the convergence of the second-order nonlinear process (6.3) is not straightforward. However, similarly to the matrix case, Theorem 4.1 allows us to show that the convergence is guaranteed for any choice of the tensor \( T \), of the matrix \( M \), and of the starting point \( y_0 \geq 0 \), provided that graph \( G_M \) is aperiodic.

**Corollary 6.1.** Let \( M : \mathbb{R}^n \to \mathbb{R}^n \) be as in Definition 3.1 and let \( G_M \) be its adjacency graph, as per Definition 4.1. If \( G_M \) is aperiodic and \( y_0 \geq 0 \), then the process \( \{y_k\}_k \) defined in (6.3) for a given seed \( \ell \) converges to a unique stationary point \( y^{(\ell)} \geq 0 \).

**Proof.** Let \( F : \mathbb{R}^n \to \mathbb{R}^n \) be the map \( F(y) = cM(y) + (1 - c)\|y\|_1 1_\ell \), where we have omitted the dependency of the map on \( \ell \) for the sake of simplicity. Note that the limit points of (6.3) coincide with the fixed points of \( F \) on the unit sphere \( \|y\|_1 = 1 \). Note moreover that \( F \) is homogeneous,
Figure 9. Link prediction performance comparison on two network dataset: UK FACULTY dataset (top) and SMALL WORLD CITATION network (bottom). The plots show means and quartiles of the ratio between the fraction of correctly predicted edges using $S_M$ and the one obtained using $S_{PR}$, over ten random trials for different values of $p$ and $\alpha$ in (3.2).

\[ F(\theta y) = \theta F(y) \], for all $\theta > 0$. Finally, notice that the $j$-th column of the Jacobian matrix of $F$ evaluated at $z$ is

\[ \frac{\partial}{\partial y_j} F(z) = c \frac{\partial}{\partial y_j} M(z) + (1 - c) 1_\ell, \]

which shows that if the Jacobian matrix of $M$ is irreducible, the same holds for the one of $F$. With these observations, the thesis follows straightforwardly using the same arguments as in the proof of Theorem 4.1, applied to $F$.

As for the linear dynamical process, the stationary distributions of (6.3) computed for different seeds allow us to define the similarity matrix $S_M$:

\[ (S_M)_{ij} = (y^{(i)})_j + (y^{(j)})_i. \]

In Figure 9 we compare the performance of the link prediction algorithm based on the standard seeded PageRank similarity matrix $S_{PR}$ (6.1) and the newly introduced similarity matrix $S_M$ (6.3) induced by $M$ with $M = P$ and $T = T_W$, the random walk triangle tensor. The tests were performed on the real-world networks UK FACULTY and SMALL WORLD CITATION. The network UK FACULTY [42] represents the personal friendships network between the faculty members of a UK University. It contains $n = 81$ vertices and $m = 817$ edges.

Please check these as the data that I’ve found refers to this as a weighted digraph. Did you symmetrize/make everything binary?

The network SMALL WORLD citation ......

Couldn’t find this network!

The experiments were performed as follows. We start with an initial network $G = (V, E)$ and we randomly select a subset of its edges, which we call $E_1$, of size $|E_1| \approx |E|/10$. We then define $G_0 = (V, E_0)$ to be the graph obtained from $G$ after removal of the edges in $E_1$, so that $E_0 = E \setminus E_1$. Thus, working on the adjacency matrix of $G_0$, we build the two similarity matrices $S_{PR}$
and $S_M$. Then, given similarity matrix $S$, we select from $V \times V \setminus E_0$ the subset $E_S$ containing the $|E_1|$ edges with the largest similarity scores $S_{ij}$. A better performance corresponds to a larger size of $E_1 \cap E_S$, since this is equivalent to detecting more of the edges that were originally in the graph. To evaluate the performance of the two similarity matrices, we thus computed the ratio $|E_S \cap E_1|/|E_{PR} \cap E_1|$. In Figure 9 we boxplot this quantity over 10 random runs where $E_1$ is sampled from the initial $E$ with a uniform probability. Whenever the boxplot is above the threshold of 1, our method is outperforming standard seeded PageRank. The middle plots in the figure display the results for the two networks when $\alpha = 0.5$ in (3.2) and we let $p$ vary. On the other hand, the plots on the right display results for varying values of $\alpha$ and $p = 0$, which was observed to achieve the best performance in the previous test. Overall, the plots clearly show that the link prediction algorithm based on the similarity matrix $S_M$ typically outperforms the one based on $S_{PR}$, especially for small values of $p.

DJH ◊ I changed “below” to “above” in the description of the box plots!

7. Conclusion

After associating a network with its adjacency matrix, it is a natural step to formulate eigenvalue problems that quantify nodal characteristics. In this work we showed that cubic tensors can be used to create a corresponding set of nonlinear eigenvalue problems that build in higher order effects; notably triangle-based motifs. Such spectral measures automatically incorporate the mutually reinforcing nature of eigenvector and PageRank centrality. As a special case, we specified a mutually reinforcing version of the classical Watts–Strogatz clustering coefficient. We showed that our general framework includes a range of approaches for combining first- and second-order interactions, and, for all of these, we gave existence and uniqueness results along with an effective computational algorithm. Synthetic and real networks were used to illustrate the approach.

Given the recent growth in activity around higher-order network features [6,7,26,27,43,44], there are many interesting directions in which this work could be further developed, including the design of such centrality measures for weighted, directed and dynamic networks, and the study of mechanistic network growth models that incorporate higher-order information.

Ethics. Insert ethics text here.

Data Accessibility. This article has no additional data.

Authors’ Contributions. The three authors contributed equally to the manuscript.

Competing Interests. Insert competing interests text here.

Funding. The work of FA was supported by fellowship ECF-2018-453 from the Leverhulme Trust. The work of DJH was supported by EPSRC/RCUK Established Career Fellowship EP/M00158X/1 and by EPSRC Programme Grant EP/P020720/1.

Acknowledgements. Insert acknowledgment text here.

Disclaimer. Insert disclaimer text here.

References


21. LaFond T, Neville J, Gallagher B. 2014 Anomaly detection in networks with changing trends. In *Outlier Detection and Description under Data Diversity at the International Conference on Knowledge Discovery and Data Mining*.


44. Yin H, Benson AR, Leskovec J, Gleich DF. 2017 Local higher-order graph clustering. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* pp. 555–564. ACM.