NON-BACKTRACKING ALTERNATING WALKS*

FRANCESCA ARRIGO†, DESMOND J. HIGHAM†, AND VANNI NOFERINI‡

Abstract. The combinatorics of walks on a graph is a key topic in network science. Here we study a special class of walks on directed graphs. We combine two features that have previously been considered in isolation. We consider *alternating walks*, which form the basis of algorithms for hub/authority detection and for discovering directed bipartite substructure. Within this class, we restrict to *non-backtracking walks*, since this constraint has been seen to offer advantages in related contexts. We derive a recursive formula for counting the total number of non-backtracking alternating walks of a given length, leading to an expression for any associated power series expansion. We discuss computational issues for the widely used cases of resolvent and exponential series, showing that non-backtracking can be incorporated at very little extra cost. We also derive an appropriate asymptotic limit which gives a parameter-free, spectral analogue. We perform tests on an artificial data set in order to quantify the advantages of the new methodology. We also show that the removal of backtracking allows us to identify larger bipartite subgraphs within an anatomical connectivity network from neuroscience.

Key words. bipartivity, centrality, directed graph, generating function, matrix polynomial, network

AMS subject classifications. 65F60, 68R10

1. Motivation.

The notion of a walk around a graph is both natural and useful. However, in some settings, walks that *backtrack*—setting out from a node and then returning to it on the next step—are best avoided. The idea of restricting attention to *non-backtracking walks* has been suggested and analysed in a wide range of fields, including spectral graph theory [2, 20, 21], number theory [34], discrete mathematics [10, 30], stochastic analysis [1], applied linear algebra [31] and computer science [29, 35]. In particular, in the area of network science non-backtracking walks have been shown to form the basis of effective algorithms for finding communities [22, 24] and assigning centrality values to nodes [6, 16, 25, 27]. In this work, we extend the non-backtracking idea to the case of *alternating walks* on directed graphs. This allows us to develop new theoretical results that lead to novel extensions of existing algorithms.

The work is organised as follows. In Section 2 we set up the basic notation and recall the idea of an alternating walk. In Sections 3 and 4 we describe how alternating walks have been used to define algorithms for discovering bipartivity and assigning nodal centrality. In Section 5 we consider non-backtracking analogues of alternating walks. After formalizing the definitions, we derive new, explicit recurrences that count the total number of non-backtracking alternating walks of any length. This gives a natural extension of the two and three term recurrences that arise in the non-alternating case. We then present a two-by-two block matrix formulation of the recurrence which allows us to deal directly with power series expansions over the walk count lengths. In this way we are able to derive computable expressions for the generalized matrix functions that arise when we define new, non-backtracking versions of the centrality and bipartivity algorithms. For traditional walk-counting centrality

*Submitted to the editors DATE.

Funding: DJH and FA are supported by grant EP/M00158X/1 from the EPSRC/RCUK Digital Economy Programme.

†Department of Mathematics and Statistics, University of Strathclyde, Glasgow, UK (francesca.arrigo@strath.ac.uk, d.j.higham@strath.ac.uk).

‡Department of Mathematical Sciences, University of Essex, Wivenhoe Park, CO4 3SQ, Colchester, UK (vnofer@essex.co.uk).

measures it is known that related, parameter-free, spectral measures can be recovered by taking an appropriate asymptotic limit. In Section 6 we use this approach to derive a new non-backtracking spectral centrality measure. Section 7 gives results on a specially constructed class of networks in order to quantify the benefits of non-backtracking. In Section 8 we consider the issue of finding directed bipartite structure in a worm brain network, and show how the non-backtracking version improves on previous results. Section 9 gives a brief summary.

2. Background and Notation. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a directed graph with n nodes and m edges. We assume that the graph is unweighted, with no self-loops or multiple edges. The associated adjacency matrix A is therefore unsymmetric and binary with zeros on the main diagonal. We let I denote the identity matrix, $\mathbf{e}_i$ the ith column of I, $\mathbf{1}$ the vector of all ones and $\mathbf{0}$ the vector of all zeros. To avoid confusion, the dimensions of these objects will sometimes be indicated with a subscript, as in I_n. Given a matrix $B \in \mathbb{R}^{n\times n}$, we use $\mathrm{diag}(B)$ to denote the diagonal matrix in $\mathbb{R}^{n\times n}$ with (i,i) element equal to $(B)_{ii}$. We denote by $\rho(B)$ the spectral radius of a matrix B and by $\sigma_1(B)$ its largest singular value.

A sequence of $k+1$ nodes $i_1, i_2, \ldots, i_{k+1}$ such that $i_\ell \to i_{\ell+1} \in \mathcal{E}$ for all $\ell = 1, 2, \ldots, k$ is called a *walk of length* k. Note that the nodes and edges are not required to be distinct. A well known result in graph theory states that the (i,j) entry of the matrix A^k counts the number of different walks of length k that start at node i and finish at node j. In directed networks, thanks to the orientation of the edges, it is possible to define *alternating walks* [7, 12, 32]. An *alternating walk of length k starting with an out-edge* is a sequence of $k+1$ nodes $i_1, i_2, \ldots, i_{k+1}$ such that $i_{\ell+1} \to i_\ell$ when ℓ is even and $i_\ell \to i_{\ell+1}$ when ℓ is odd. Walks of this type will thus have the form:

$$i_1 \to i_2 \leftarrow \cdots \to i_k \leftarrow i_{k+1}, \quad \text{if } k \text{ is even}$$

and

$$i_1 \to i_2 \leftarrow \cdots \leftarrow i_k \to i_{k+1}, \quad \text{if } k \text{ is odd.}$$

It is easily seen that the kth power of AA^T contains in its (i,j) entry the number of alternating walks from node i to node j of length $2k$ starting with an out-edge (note that, since $2k$ is even, these walks will end with an edge of the form $\leftarrow$). In a similar way, entries of the matrix $(AA^T)^k A$ count alternating walks of length $2k+1$ starting with an out-edge and ending with $\to$. We may analogously define an *alternating walk of length k starting with an in-edge* as a sequence of $k+1$ nodes $i_1, i_2, \ldots, i_{k+1}$ such that

$$i_1 \leftarrow i_2 \to \cdots \leftarrow i_k \to i_{k+1}, \quad \text{if } k \text{ is even}$$

and

$$i_1 \leftarrow i_2 \to \cdots \to i_k \leftarrow i_{k+1}, \quad \text{if } k \text{ is odd,}$$

i.e., such that for all $\ell = 1, 2, \ldots, k$ we have $i_\ell \to i_{\ell+1}$ when ℓ is even, and $i_{\ell+1} \to i_\ell$ when ℓ is odd. The number of alternating walks of length $2k$ starting with an in-edge (and therefore ending with $\to$) is counted by $(A^TA)^k$, and $(A^TA)^kA^T$ counts the number of alternating walks of length $2k+1$ starting with an in-edge (and therefore ending with $\leftarrow$).

Within the field of network science, the concept of alternating walks has been used in two related areas, as we discuss in the next two sections.

3. Detecting Bipartite Structures. The use of alternating walk counts in network science first appeared in [12], where the authors considered the problem of uncovering directed bipartite subnetworks. More precisely, they wanted to find two disjoint sets of nodes $S_1, S_2 \subset \mathcal{V}$ such that:

- nodes within each set have few links between them, and
- there are several links from S_1 to S_2 and few from S_2 to S_1.

The existence of this type of substructure reveals that nodes in S_1, although not strongly interconnected, share a common topological role in the network—they form a one-way bipartite route into the nodes of S_2. Similarly, the nodes in S_2 are not strongly interconnected, but share the common role of endpoints for these edges. In [12] this type of pattern was seen to be of relevance in an anatomical neural connectivity network, where discovering a subset S_1 was consistent with identifying "command neurons" that pass messages down the hierarchy.

The algorithm developed in [12] made use of the matrix

$$F(A) = \left(I + \frac{AA^T}{2!} + \frac{(AA^T)^2}{4!} + \cdots\right) - \left(A + \frac{AA^TA}{3!} + \frac{A(A^TA)^2}{5!} + \cdots\right). \tag{3.1}$$

The expression in the first pair of parentheses counts alternating walks of even length starting with an out-edge, while the second counts the number of alternating walks of odd length starting with an out-edge. If the required bipartite structure existed in the directed graph, we would have many of these alternating walks of even length that start and finish in S_1. Likewise, we would have many alternating walks of odd length that start in S_1 and finish in S_2. Based on this interpretation of $F(A)$, one would expect $F(A)_{ij}$ to take large positive values when $i, j \in S_1$ and large negative values when $i \in S_1$ and $j \in S_2$. Similarly, $(F(A^T))_{ij}$ will take large positive values when both nodes belong to S_2 and large negative values when $i \in S_2$ and $j \in S_1$. Therefore, $F(A)+F(A^T)$ is expected to reveal intra-cluster ($S_1 \to S_1$ and $S_2 \to S_2$) relationships through positive entries and inter-cluster ($S_1 \to S_2$ or $S_2 \to S_1$) relationships through negative entries. A very useful feature of this approach is that the task is reduced to finding strongly connected clusters in the symmetric, weighted, network represented by $F(A) + F(A^T)$. This has converted the problem to a standard form where many well-tested algorithms are available.

For our purposes, it is convenient to write $F(A)$ in (3.1) as

$$F(A) = \cosh(\sqrt{AA^T}) - \sinh^{\diamond}(A).$$

Here, given a compact SVD of the rank-r matrix $A = U_r\Sigma_r V_r^T$, we let $f^{\diamond}(A) = U_r f(\Sigma_r) V_r^T$ be the generalized matrix function of A induced by f [4, 17]. More generally, the definition of $F(A)$ and its interpretation can be extended to any matrix function $f(x) = \sum_{k=0}^{\infty} c_k x^k$ defined on the spectrum of AA^T, with even and odd parts

$$f_{\text{even}}(x) = \frac{f(x) + f(-x)}{2}, \quad \text{and} \quad f_{\text{odd}}(x) = \frac{f(x) - f(-x)}{2},$$

by forming

$$\mathcal{F}_f(A) = f_{\text{even}}(\sqrt{AA^T}) - f^{\diamond}_{\text{odd}}(A). \tag{3.2}$$

To see this, we note that $f_{\text{even}}(\sqrt{AA^T}) = \sum_k c_{2k}(AA^T)^k$ has in its (i,j) entry a weighted sum of all alternating walks of even length starting at node i with an out-edge and ending at node j. Similarly, $f^{\diamond}_{\text{odd}}(A) = \sum_k c_{2k+1}(AA^T)^k A$ contains a weighted sum of all alternating walks of odd length starting with an out-edge.

Remark 3.1. Odd generalized matrix functions of a matrix A can be expressed via standard matrix functions of the matrix AA^T or A^TA; see, e.g., [4, 17].

The original definition (3.1) is based on an exponential function expansion. In this work we will also consider an expansion of the resolvent function, $f(x) = (1-\alpha x)^{-1}$ with $\alpha \in (0, 1/\sigma_1(A))$, with $\sigma_1(A)$ denoting the largest singular value of A. We therefore define the matrix

$$G(A) = (I - \alpha^2 AA^T)^{-1} - g^\diamond(A), \tag{3.3}$$

where $g(x) = \alpha x/(1-\alpha^2 x^2)$ with $\alpha \in (0, 1/\sigma_1(A))$. In summary, the matrices $F(A)+F(A^T)$ and $G(A)+G(A^T)$ will be used for revealing intra-cluster relationships through positive entries and inter-cluster relations through negative entries.

4. Centrality Measures. Nodes in a directed network play two roles, both spreading and receiving information, and thus two types of centrality are relevant. Kleinberg [23] quantified these centralities through the concepts of *hub* and *authority* measures based on the HITS algorithm. Intuitively, a good hub points to many good authorities and a good authority is pointed to by many good hubs. This defines a recursive relationship between the two centrality measures. The HITS algorithm uses this idea to compute a nonnegative vector of hub centralities, $\mathbf{x}^\star$, and authority centralities, $\mathbf{y}^\star$. In each case the ith component of the vector represents the centrality of node i, with a larger value indicating greater centrality. Given two starting vectors $\mathbf{x}^{(0)} = \mathbf{y}^{(0)} = \mathbf{1}/\sqrt{n}$, the algorithm iterates for $k = 1, 2, \dots$ until convergence

$$\begin{cases} \mathbf{x}^{(k)} = A\mathbf{y}^{(k-1)}, \\ \mathbf{y}^{(k)} = A^T\mathbf{x}^{(k)}, \end{cases}$$

followed by a normalization step. By substitution it follows that

$$\begin{cases} \mathbf{x}^{(k)} = AA^T\mathbf{x}^{(k-1)} = (AA^T)^k\mathbf{x}^{(0)}, \\ \mathbf{y}^{(k)} = A^TA\mathbf{y}^{(k-1)} = (A^TA)^k\mathbf{y}^{(0)}, \end{cases}$$

followed by normalization. Hence, HITS is a power method to compute the eigenvectors associated with the leading eigenvalues of AA^T, the *hub matrix*, and A^TA, the *authority matrix*, i.e., the first left and right singular vectors of A [19].

We saw in Section 2 that powers of the hub and authority matrices can be used to count alternating walks in directed networks. Hence, HITS may also be interpreted from this perspective: a good hub is pointed to by many good authorities which are themselves pointed to by many good hubs, and so on. So a good hub initiates many alternating walks around the network that start with an out-edge. Similarly, a good authority initiates many alternating walks around the network that start with an in-edge. Following on from HITS, further centrality measures have been proposed as entries of functions of the hub and authority matrices or row and columns sums of these; see, [3, 7].

Using the general notation of Section 3, where $f(x) = \sum_{k=0}^{\infty} c_k x^k$ is a function defined on the spectrum of AA^T and letting $f_{\text{even}}(x)$ and $f_{\text{odd}}(x)$ denote its even and odd parts, we see that, within the radius of convergence, the vector $\mathbf{h} = (h_i)$ whose components are

$$h_i = \mathbf{e}_i^T f_{\text{even}}(\sqrt{AA^T})\mathbf{z}_i +^T \mathbf{e}_i^T f_{\text{odd}}^\diamond(A)\mathbf{w}_i$$

is a candidate for measuring *hub centrality*, and the vector $\mathbf{a} = (a_i)$ having components

$$a_i = \mathbf{e}_i^T f_{\text{even}}(\sqrt{A^T A})\mathbf{z}_i + \mathbf{e}_i^T f_{\text{odd}}^{\diamond}(A^T)\mathbf{w}_i$$

is a candidate for measuring *authority centrality*. Here, $\mathbf{z}_i, \mathbf{w}_i \in \{\mathbf{e}_i, \mathbf{1}, \mathbf{0}\}$ for all $i = 1, 2, \ldots, n$ with the constraint that the same "type" of vector should be selected for all values of i. Namely, if for example we select $\mathbf{z}_{i^*} = \mathbf{e}_{i^*}$ and $\mathbf{w}_{i^*} = \mathbf{0}$ for some i^*, then we should select $\mathbf{z}_i = \mathbf{e}_i$ and $\mathbf{w}_i = \mathbf{0}$ for all $i = 1, 2, \ldots, n$.

Before moving on to the description of non-backtracking alternating walks, let us briefly discuss the centrality measures resulting from different common choices of the vectors $\mathbf{z}_i$ and $\mathbf{w}_i$. One of the most popular choices is represented by $\mathbf{z}_i = \mathbf{e}_i$ and $\mathbf{w}_i = \mathbf{0}$ for all i, or vice versa; with this choice of the vectors the resulting centrality measures are *f*-subgraph centrality type of measures [14]. In particular, if $f(x) = e^x$ and $\mathbf{z}_i = \mathbf{e}_i$ and $\mathbf{w}_i = \mathbf{0}$, we derive the hub and authority centrality measures introduced in [7]. On the other hand, using the same function but taking $\mathbf{z}_i = \mathbf{0}$ and $\mathbf{w}_i = \mathbf{e}_i$ yields the hub-authority centrality measure; see [7]. The choice of $\mathbf{z}_i = \mathbf{0}$ and $\mathbf{w}_i = \mathbf{1}$ for all i results in a total (node) communicability type of measure [8]. When considering $f(x) = (1 - \alpha x)^{-1}$ for appropriate choices of $\alpha > 0$ we obtain the total communicability of nodes introduced in [3]. Another choice of vectors that may lead to new, insightful centrality measures is $\mathbf{z}_i = \mathbf{w}_i = \mathbf{1}$. This yields again to a total (node) communicability type of centrality measure, where each node is targeting all the nodes in a network, regardless of whether they are acting as hubs or authorities in the graph.

5. Non-Backtracking Alternating Walks. In Sections 3 and 4 we described how the combinatorics of alternating walks on directed graphs are useful for addressing issues in bipartivity and centrality. In both cases, we intuitively rely on the fact that the walks are exploring the network—by taking account of walks of length greater than one we build in global information. However, with this methodology we give equal weight to all walks of the same length. Ideally, for a given walk length, we would like to give more weight to walks that roam further from their starting point. A very effective compromise that retains the efficient linear algebra underlying walk-counting while biasing the count towards more desirable walks arises when we eliminate backtracking. A walk is said to backtrack if, at any stage, it takes an edge out of a node and then immediately takes the reciprocal edge back into it. For standard walks, this idea was studied in [16] for undirected networks and in [5, 6] in the directed case. Because the walk-counting measures correspond to spectral measures as we approach the radius of convergence of the power series, related measures that use the Hashimoto or non-backtracking matrix are also relevant [24, 25].

From [5, 6, 16, 24, 25, 29] we know that advantages of the non-backtracking approach include

- low computational cost—comparable with, or less than, the backtracking counterparts,
- avoidance of localization (where most of the measure is assigned to a finite subset of the nodes),
- greater flexibility—a larger radius of convergence in the power series, which leads to a wider choice of attenuation parameter (corresponding to α in (3.3)).

This motivates our interest in defining and analyzing non-backtracking analogues of alternating walks for use in centrality and bipartivity algorithms.

Formally, we will say that an alternating walk is *backtracking* if it contains a sequence of the form $i \to j \leftarrow i$ or $i \leftarrow j \to i$, otherwise we will say that it is *non-*

backtracking. For the sake of brevity, we will use the acronyms BTAW and NBTAW for the phrases backtracking alternating walk and non-backtracking alternating walk, respectively.

We let $p_k(A)$ denote the matrix whose (i,j) entry records the number of NBTAWs from node i to node j of length k that start with an out-edge. Similarly, we let $q_k(A)$ record the number of NBTAWs from node i to node j of length k that start with an in-edge. Notice that, by construction, we have

$$p_k(A^T) = q_k(A), \quad \text{for all } k, A. \tag{5.1}$$

Our aim is to derive computable expressions for the associated power series

$$\phi(A, \{c_k\}) = \sum_{k=0}^{\infty} c_k p_k(A)$$

and

$$\varphi(A, \{c_k\}) = \sum_{k=0}^{\infty} c_k q_k(A),$$

as well as their even and odd parts. We note in passing that from (5.1) it follows that $\phi(A^T, \{c_k\}) = \varphi(A, \{c_k\})$.

Using these series, new centrality vectors can be computed as $\widehat{\mathbf{h}} = (\widehat{h}_i)$ with

$$\widehat{h}_i = \mathbf{e}_i^T \phi_{\text{even}}(A, \{c_k\})\mathbf{z}_i + \mathbf{e}_i^T \phi_{\text{odd}}(A, \{c_k\})\mathbf{w}_i \tag{5.3a}$$

and $\widehat{\mathbf{a}} = (\widehat{a}_i)$ with

$$\widehat{a}_i = \mathbf{e}_i^T \varphi_{\text{even}}(A, \{c_k\})\mathbf{z}_i + \mathbf{e}_i^T \varphi_{\text{odd}}(A, \{c_k\})\mathbf{w}_i, \tag{5.3b}$$

for some $\mathbf{z}_i, \mathbf{w}_i \in \{\mathbf{1}, \mathbf{0}, \mathbf{e}_i\}$ for all $i = 1, 2, \dots, n$. Similarly, a non-backtracking analogue of $\mathcal{F}_f(A)$ takes the form

$$\widehat{\mathcal{F}}_f(A) = \phi_{\text{even}}(A, \{c_k\}) - \phi_{\text{odd}}(A, \{c_k\}), \tag{5.4}$$

that is,

$$\widehat{\mathcal{F}}_f(A) = (c_0 I + c_2 p_2(A) + c_4 p_4(A) + \cdots) - (c_1 A + c_3 p_3(A) + \cdots).$$

We emphasize that this is the non-backtracking analogue of the walk-counting expression (3.2), and hence $\widehat{\mathcal{F}}_f(A) + \widehat{\mathcal{F}}_f(A^T)$ is a corresponding indicator for directed bipartite structures.

5.1. Recurrence Relations. The following result gives a two-term recurrence for the matrices $p_k(A)$ and $q_k(A)$. This result may be compared with the corresponding recurrence that applies for standard walks on directed graphs [10], see also [31].

THEOREM 5.1. *Let A be the adjacency matrix of a directed, unweighted graph with no self-loops or multiple edges. Then, in the above notation,*

$$p_1(A) = A, \quad p_2(A) = AA^T - D_1,$$

where $D_1 = D_1(A) = \operatorname{diag}(AA^T)$, and

$$q_1(A) = A^T, \quad q_2(A) = A^T A - D_2,$$

where $D_2 = D_2(A) = \operatorname{diag}(A^T A)$. Moreover, for $k \geq 2$

- *if k is even, then*

$$p_{k+1}(A) = p_k(A)A + p_{k-1}(A)(I - D_2)$$

and

$$q_{k+1}(A) = q_k(A)A^T + q_{k-1}(A)(I - D_1),$$

- *if k is odd, then*

$$p_{k+1}(A) = p_k(A)A^T + p_{k-1}(A)(I - D_1)$$

and

$$q_{k+1}(A) = q_k(A)A + q_{k-1}(A)(I - D_2).$$

Proof. We proceed by induction. It is straightforward to check the base case concerning $p_1(A), p_2(A), q_1(A)$ and $q_2(A)$. For $k \geq 2$ and k even, we assume as our inductive hypothesis that all the matrices up to $p_k(A)$ properly count NBTAWs. We will show that $p_k(A)A + p_{k-1}(A)(I - D_2)$ produces $p_{k+1}(A)$. Similar arguments can be used for the other iterations stated in the theorem.

Let $i, j \in \mathcal{V}$ be two nodes in the network. Any NBTAW of length $k+1$ starting at i with an out-edge and ending at j can be obtained by propagating NBTAWs of length k starting at i with an out-edge and arriving at some neighbour $\ell \in \mathcal{V}$ of j such that $\ell \to j$. The total of all such walks, which have the form

$$i \to \cdots \leftarrow \ell \to j \qquad \text{of length } k+1,$$

is counted by $(p_k(A)A)_{ij}$. However, this total includes walks of the form

$$\overbrace{\underbrace{i \to \cdots j \leftarrow \ell}_{\text{NBTAW}} \to j}^{\text{BTAW}} \qquad \text{of length } k+1$$

that are backtracking, and thus need to be removed from the count. Because these unwanted walks take the form of a NBTAW of length $(k-1)$ followed by $j \leftarrow \ell \to j$, it is natural to deal with this issue by subtracting $(p_{k-1}(A)D_2)_{ij}$. However, subtracting this quantity will also remove from the count walks of the form

$$\overbrace{\underbrace{i \to \cdots \leftarrow \ell \to j}_{\text{NBTAW}} \leftarrow \ell \to j}^{\text{BTAW}} \qquad \text{of length } k+1$$

which have already been taken care of at an earlier stage. In order to compensate we add the quantity $(p_{k-1}(A))_{ij}$, since the walks that need to be removed are in a one-to-one relationship with walks of the form

$$\underbrace{i \to \cdots \leftarrow \ell \to j}_{\text{NBTAW}} \qquad \text{of length } k-1.$$

This concludes the proof. ∎

Remark 5.2. It is interesting to note that in the case of standard walks, the analogous, simpler, recurrences that correspond to those in Theorem 5.1 change from having two terms to three terms when we move from undirected to directed graphs [10, 30, 31]. Theorem 5.1 applies to directed graphs and yet involves only two-term recurrences. Intuitively, this discrepancy arises because in the case of alternating walks every directed edge offers an opportunity to track back immediately to the previous node—in this sense, directed edges in the world of alternating walks are analogous to undirected edges in the world of standard walks.

Our next result shows that the pair of recurrences in Theorem 5.1 can be reformulated as a single recurrence in a higher dimension.

THEOREM 5.3. *Let*

$$\mathcal{A} = \begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix} \in \mathbb{R}^{2n\times 2n}. \tag{5.5}$$

Now let $r_k(\mathcal{A})$ *be defined by*

$$r_1(\mathcal{A}) = \mathcal{A}, \qquad r_2(\mathcal{A}) = \mathcal{A}^2 - \Delta,$$

where $\Delta = \operatorname{diag}(\mathcal{A}^2)$*, and, for* $k \geq 2$*,*

$$r_{k+1}(\mathcal{A}) = r_k(\mathcal{A})\mathcal{A} + r_{k-1}(\mathcal{A})(I - \Delta). \tag{5.6}$$

Then,

$$r_{2k}(\mathcal{A}) = \begin{bmatrix} p_{2k}(A) & 0 \\ 0 & q_{2k}(A) \end{bmatrix} \quad \textit{and} \quad r_{2k+1}(\mathcal{A}) = \begin{bmatrix} 0 & p_{2k+1}(A) \\ q_{2k+1}(A) & 0 \end{bmatrix}.$$

Proof. From the structure of $\mathcal{A}$ it follows that

$$\Delta = \begin{bmatrix} D_1 & 0 \\ 0 & D_2 \end{bmatrix},$$

where D_1 and D_2 are defined in the statement of Theorem 5.1. It is straightforward to check that the result holds for $r_1(\mathcal{A})$ and $r_2(\mathcal{A})$. Assume the result is true up to some level k. Then, if k is even

$$\begin{aligned} r_{k+1}(\mathcal{A}) &= \begin{bmatrix} p_k(A) & 0 \\ 0 & q_k(A) \end{bmatrix} \begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix} \\ &\quad + \begin{bmatrix} 0 & p_{k-1}(A) \\ q_{k-1}(A) & 0 \end{bmatrix} \begin{bmatrix} I - D_1 & 0 \\ 0 & I - D_2 \end{bmatrix} \\ &= \begin{bmatrix} 0 & p_k(A)A + p_{k-1}(A)(I - D_2) \\ q_k(A)A^T + q_{k-1}(A)(I - D_1) & 0 \end{bmatrix}. \end{aligned}$$

Similarly, if k is odd then

$$r_{k+1}(\mathcal{A}) = \begin{bmatrix} p_k(A)A^T + p_{k-1}(A)(I - D_1) & 0 \\ 0 & q_k(A)A + q_{k-1}(A)(I - D_2) \end{bmatrix}.$$

Hence, by Theorem 5.1, the result follows. ∎

Remark 5.4. The use of the block matrix $\mathcal{A}$ in (5.5) is intimately connected with an equivalence between directed graphs and undirected bipartite graphs. This equivalence has previously been exploited in [7, 11]. Powering up the matrix $\mathcal{A}$ counts standard walks in an undirected bipartite graph, which corresponds to counting alternating walks in the original digraph. In a similar manner, the recurrence in Theorem 5.3 counts non-backtracking walks in an undirected bipartite graph, which corresponds to counting NBTAWs in the original directed graph. The compact formulation in Theorem 5.3 will allow us access to convenient expressions for the required power series expansions, using generating function techniques.

It follows from Theorem 5.3 that $\phi_{\text{even}}(A, \{c_k\})$, $\phi_{\text{odd}}(A, \{c_k\})$, $\varphi_{\text{even}}(A, \{c_k\})$ and $\varphi_{\text{odd}}(A, \{c_k\})$ in (5.3a) and(5.3b) may be computed via

$$\begin{aligned}\sum_{k=0}^{\infty} c_k r_k(\mathcal{A}) &= \begin{bmatrix} \sum_k c_{2k} p_{2k}(A) & \sum_k c_{2k+1} p_{2k+1}(A) \\ \sum_k c_{2k+1} q_{2k+1}(A) & \sum_k c_{2k} q_{2k}(A) \end{bmatrix} \\ &= \begin{bmatrix} \phi_{\text{even}}(A, \{c_k\}) & \phi_{\text{odd}}(A, \{c_k\}) \\ \varphi_{\text{odd}}(A, \{c_k\}) & \varphi_{\text{even}}(A, \{c_k\}) \end{bmatrix}.\end{aligned}$$

Hence, the non-backtracking centrality measures $\widehat{\mathbf{h}}$ and $\widehat{\mathbf{a}}$ may be written as

$$\widehat{h}_i = [\mathbf{e}_i^T, \mathbf{0}^T] \left(\sum_{k=0}^{\infty} c_k r_k(\mathcal{A}) \right) \begin{bmatrix} \mathbf{z}_i \\ \mathbf{w}_i \end{bmatrix} \tag{5.7}$$

and

$$\widehat{a}_i = [\mathbf{0}^T, \mathbf{e}_i^T] \left(\sum_{k=0}^{\infty} c_k r_k(\mathcal{A}) \right) \begin{bmatrix} \mathbf{w}_i \\ \mathbf{z}_i \end{bmatrix}. \tag{5.8}$$

Similarly, (5.4) rewrites as

$$\widehat{\mathcal{F}}_f(A) = [I, 0] \left(\sum_{k=0}^{\infty} c_k r_k(\mathcal{A}) \right) \begin{bmatrix} I \\ -I \end{bmatrix}. \tag{5.9}$$

These expressions may be simplified with the help of the following lemma.

LEMMA 5.5. *Writing* $f_h(x) = \sum_{k=0}^{\infty} c_{k+h} x^k$, *we have*

$$\sum_{k=0}^{\infty} c_k r_k(\mathcal{A}) = [I_{2n}, 0] \left(f_0(Y) - f_2(Y) \right) \begin{bmatrix} I_{2n} \\ 0 \end{bmatrix},$$

where

$$Y = \begin{bmatrix} \mathcal{A} & I - \Delta \\ I & 0 \end{bmatrix}, \tag{5.10}$$

whenever the series converge.

Proof. The recurrence (5.6) with A instead of $\mathcal{A}$ was studied in [5] using techniques from the theory of generating functions. Exactly the same arguments may be used in this case. ∎

From Lemma 5.5, we see that $\widehat{\mathbf{h}}$ in (5.7), $\widehat{\mathbf{a}}$ in (5.8) and $\widehat{\mathcal{F}}_f(A)$ in (5.9) may be computed in terms of the matrix functions $f_0(Y)$ and $f_2(Y)$. We now examine the two most popular cases, and show that convenient formulas are available.

5.2. Exponential. In the case where $c_k = t^k/k!$ for some fixed t, so that $f(x) = e^{tx}$, it follows from (5.7), (5.8) and Lemma 5.5 that

$$\widehat{h}_i = [\mathbf{0}^T, \mathbf{0}^T, \mathbf{e}_i^T, \mathbf{0}^T] t\psi_1(tY) \begin{bmatrix} (\mathcal{A}^2 - \Delta)\mathbf{z}_i \\ \mathcal{A}\mathbf{w}_i \end{bmatrix} + \mathbf{e}_i^T \mathbf{z}_i$$

and

$$\widehat{a}_i = [\mathbf{0}^T, \mathbf{0}^T, \mathbf{0}^T, \mathbf{e}_i^T] t\psi_1(tY) \begin{bmatrix} (\mathcal{A}^2 - \Delta)\mathbf{w}_i \\ \mathcal{A}\mathbf{z}_i \end{bmatrix} + \mathbf{e}_i^T \mathbf{z}_i.$$

(This may be compared with the result for standard non-backtracking walks in [5].) Moreover, in (5.9) we obtain

$$\begin{aligned} \widehat{F}(A) &= \phi_{\text{even}}(A, \{1/k!\}) - \phi_{\text{odd}}(A, \{1/k!\}) \\ &= [I, 0] \left(\sum_{k=0}^{\infty} \frac{p_k(\mathcal{A})}{k!} \right) \begin{bmatrix} I \\ -I \end{bmatrix} \\ &= [0, 0, I, 0] \psi_1(Y) \begin{bmatrix} AA^T - D_1 \\ D_2 - A^T A \\ -A \\ A^T \end{bmatrix} + I. \end{aligned}$$

In these expressions, we need to compute $\psi_1(Y)V$, for some matrix V of appropriate size. This can be done by using the matrix

$$\widetilde{Y} = \begin{bmatrix} Y & V \\ 0 & 0 \end{bmatrix},$$

since it holds that [28]:

$$\exp(\widetilde{Y}) = \begin{bmatrix} e^Y & \psi_1(Y)V \\ 0 & I \end{bmatrix}.$$

In particular, we can compute $\widehat{F}(A)$ as

$$\widehat{F}(A) = [0, 0, I, 0, 0] \exp(\widetilde{Y}) \begin{bmatrix} 0 \\ I \end{bmatrix} + I$$

and similarly

$$\widehat{F}(A^T) = -[0, 0, 0, I, 0] \exp(\widetilde{Y}) \begin{bmatrix} 0 \\ I \end{bmatrix} + I.$$

5.3. Resolvent. We now consider $c_k = t^k$ for some $t \in (0, 1/\rho(Y))$, so $f(x) = (1 - tx)^{-1}$. In this case (5.7), (5.8) and Lemma 5.5 give

$$\left(\sum_{k=0}^{\infty} t^k r_k(\mathcal{A}) \right) M(t) = (1 - t^2) I,$$

where

$$M(t) = I - t\mathcal{A} + t^2(\Delta - I). \tag{5.11}$$

(This may be compared with the result for standard non-backtracking walks in [16].) We may thus write (5.7) and (5.8) as

$$\widehat{h}_i = (1 - t^2) [\mathbf{e}_i^T, \mathbf{0}^T] M(t)^{-1} \begin{bmatrix} \mathbf{z}_i \\ \mathbf{w}_i \end{bmatrix}$$

and

$$\widehat{a}_i = (1-t^2)\,[\mathbf{0}^T,\, \mathbf{e}_i^T]M(t)^{-1}\begin{bmatrix}\mathbf{w}_i\\ \mathbf{z}_i\end{bmatrix}.$$

Analogously, we can write $\widehat{G}(A) := \widehat{\mathcal{F}}_f(A)$ induced by $f(x) = (1-tx)^{-1}$ as

$$\widehat{G}(A) = (1-t^2)\,[I,\, 0](I - t\mathcal{A} + t^2(\Delta - I))^{-1}\begin{bmatrix} I\\ -I\end{bmatrix},$$

and similarly

$$\widehat{G}(A^T) = (1-t^2)\,[0,\, I](I - t\mathcal{A} + t^2(\Delta - I))^{-1}\begin{bmatrix} -I\\ I\end{bmatrix}.$$

5.4. Detecting bipartite subnetworks in digraph. We explained in Section 3 how the matrices $F(A) + F(A^T)$ and $G(A) + G(A^T)$ can be used for revealing intra-cluster relationships through positive entries and inter-cluster relations through negative entries. We have now shown that their non-backtracking counterparts $\widehat{F}(A) + \widehat{F}(A^T)$ and $\widehat{G}(A) + \widehat{G}(A^T)$ can be computed efficiently. Eliminating backtracking is intuitively reasonable, since it focuses attention on traversals that explore more of the network, and it has proved to be effective for related tasks [5, 6, 16, 24, 25, 29]. In Sections 7 and 8 we compare the performance of $F(A)+F(A^T)$ and $G(A) + G(A^T)$ with that of $\widehat{F}(A) + \widehat{F}(A^T)$ and $\widehat{G}(A) + \widehat{G}(A^T)$ for detecting bipartite subnetworks in digraphs.

6. Limiting behavior. Resolvent-based centrality measures have been studied in the limit as the attenuation parameter approaches the radius of convergence of the associated power series from below. The standard Katz case was analyzed in [9] and its non-backtracking analogue in [16]. These results gave useful insights by showing that the walk-based versions approach centrality measures that had been defined in the literature from a spectral perspective. In our non-backtracking alternating walk context, no such spectral measure has been previously defined, and in this section we show that the limit $t \to \mu^-$, where $\mu = 1/\rho(Y)$, gives rise to a new, parameter-free centrality measure.

We may regard $M(t)$ in (5.11) as a matrix polynomial [15], and it has the specific structure of a deformed graph Laplacian. The literature on matrix polynomials is extremely rich, see, e.g., [13, 15, 26] and the references therein. Here, we just need to recall the definition of a (right) eigenpair of a regular matrix polynomial. An $n \times n$ square matrix $P(t)$ whose entries are polynomials in t with coefficients in some field $\mathbb{F}$ is called regular if $\det P(t) \neq 0$. In this case, a finite eigenpair of $P(t)$ is a pair $(\lambda, \mathbf{v})$, such that $P(\lambda)\mathbf{v} = \mathbf{0}$, and with $\lambda \in \mathbb{K}, \mathbf{v} \in \mathbb{K}^n \setminus \{\mathbf{0}\}$, where $\mathbb{K}$ is the algebraic closure of $\mathbb{F}$. (In the case of our interest, the base field is $\mathbb{R}$ and the finite eigenvalues are sought in $\mathbb{C}$.)

Spectral properties of the deformed graph Laplacian, in relation to properties of the underlying graph, were extensively studied in [16]. Here, we are concerned with adjacency matrices of the form $\mathcal{A}$ in (5.5). We first show that this forces the spectrum of the deformed graph Laplacian to be symmetric about zero.

PROPOSITION 6.1. *It holds that $\lambda \in \mathbb{C}$ is as an eigenvalue of $M(t) = I - t\mathcal{A} + t^2(\Delta - I)$ with eigenvector $\begin{bmatrix}\mathbf{v}_1\\ \mathbf{v}_2\end{bmatrix}$ if and only if $-\lambda$ is an eigenvalue with eigenvector $\begin{bmatrix}\mathbf{v}_1\\ -\mathbf{v}_2\end{bmatrix}$.*

Proof. The statement is an immediate consequence of the identity

$$\begin{bmatrix} I_n & 0 \\ 0 & -I_n \end{bmatrix} M(t) \begin{bmatrix} I_n & 0 \\ 0 & -I_n \end{bmatrix} = M(-t)$$

and of the fact that, for any pair X, Y of invertible matrices, $(\lambda, Y\mathbf{v})$ is an eigenpair for $M(t)$ if and only if $(\lambda, \mathbf{v})$ is an eigenpair for $XM(t)Y$. ∎

In the remainder of this section we will assume $\mathcal{A}$ irreducible; this happens precisely when both AA^T and A^TA are irreducible. Note that if $\mathcal{A}$ was reducible, we could work independently on each of the adjacency matrices corresponding to the different connected components in the associated bipartite graph.

Let $\mu = 1/\rho(Y) \leq 1$ denote the smallest positive eigenvalue of the deformed graph Laplacian $M(t)$. (The set of real positive eigenvalues of $M(t)$ is not empty, since 1 is always an eigenvalue [16, Proposition 4.4.2].) Since the underlying graph is connected, it follows from [16, Proposition 7.5] that μ is simple; moreover, the corresponding eigenvector $\mathbf{v} := \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \neq \mathbf{0}$ can be taken to be componentwise nonnegative (see, [16, Theorem 10.2]). It thus holds that

$$[I - \mu\mathcal{A} + \mu^2(\Delta - I)]\mathbf{v} = \mathbf{0},$$

or equivalently,

$$(1-\mu^2)\mathbf{x} + \mu^2 D_1\mathbf{x} = \mu(A\mathbf{y}) \text{ and } (1-\mu^2)\mathbf{y} + \mu^2 D_2\mathbf{y} = \mu(A^T\mathbf{x}).$$

Hence it follows that

$$Y \begin{bmatrix} \mathbf{v} \\ \mu\mathbf{v} \end{bmatrix} = \frac{1}{\mu} \begin{bmatrix} \mathbf{v} \\ \mu\mathbf{v} \end{bmatrix}$$

with $\mu^{-1} = \rho(Y)$ and the eigenvector is again componentwise nonnegative since $\mathbf{v} \geq \mathbf{0}$.

We may now characterizes the limit of interest.

THEOREM 6.2. *Suppose that $\mathcal{A}$ is irreducible. In the above notation, there exists a constant $c > 0$ such that*

$$\lim_{t \to \mu^-} (\mu - t) M(t)^{-1} = c\, \mathbf{v}\mathbf{v}^T.$$

Proof. Following the proof of [16, Theorem 10.1], as well as the fact that μ is a simple eigenvalue of the deformed graph Laplacian [16, Proposition 7.5], we have that there exists a positive constant c such that

$$(I + t\mathcal{A} + t^2(\Delta - I))^{-1} = \frac{c}{\mu - t} \begin{bmatrix} \mathbf{x}\mathbf{x}^T & \mathbf{x}\mathbf{y}^T \\ \mathbf{y}\mathbf{x}^T & \mathbf{y}\mathbf{y}^T \end{bmatrix} + o\left(\frac{1}{\mu - t}\right).$$ ∎

Theorem 6.2 can be used to deduce the limiting behaviour of our quantities of interest, taking into account that centralities can always be renormalized (as only the ratios of the values of their entries are relevant to ranking nodes). We see that

$$\begin{aligned}(\mu - t)\left(\widehat{G}(A) + \widehat{G}(A^T)\right) &= (\mu - t)[I,\, I](I - t\mathcal{A} + t^2(\Delta - I))^{-1} \begin{bmatrix} I \\ -I \end{bmatrix} \\ &\to c(\mathbf{x} + \mathbf{y})(\mathbf{x} - \mathbf{y})^T.\end{aligned}$$

TABLE 1
Limiting behavior of the measures defined in (5.3a) and (5.3b) when $c_r = t^r$ as $t \to 1/\rho(Y)$.

$\mathbf{z}_i =$ \ $\mathbf{w}_i =$	$\mathbf{e}_i$	$\mathbf{1}$	$\mathbf{0}$
$\mathbf{e}_i$	$\widehat{\mathbf{h}} \to \mathbf{x} \circ (\mathbf{x} + \mathbf{y})$ $\widehat{\mathbf{a}} \to \mathbf{y} \circ (\mathbf{x} + \mathbf{y})$	$\widehat{\mathbf{h}} \to \|\mathbf{y}\|_1 \mathbf{x} + \mathbf{x} \circ \mathbf{x}$ $\widehat{\mathbf{a}} \to \|\mathbf{x}\|_1 \mathbf{y} + \mathbf{y} \circ \mathbf{y}$	$\widehat{\mathbf{h}} \to \mathbf{x} \circ \mathbf{x}$ $\widehat{\mathbf{a}} \to \mathbf{y} \circ \mathbf{y}$
$\mathbf{1}$	$\widehat{\mathbf{h}} \to \|\mathbf{x}\|_1 \mathbf{x} + \mathbf{x} \circ \mathbf{y}$ $\widehat{\mathbf{a}} \to \|\mathbf{y}\|_1 \mathbf{y} + \mathbf{x} \circ \mathbf{y}$	$\widehat{\mathbf{h}} \to \mathbf{x}$ $\widehat{\mathbf{a}} \to \mathbf{y}$	$\widehat{\mathbf{h}} \to \mathbf{x}$ $\widehat{\mathbf{a}} \to \mathbf{y}$
$\mathbf{0}$	$\widehat{\mathbf{h}} \to \mathbf{x} \circ \mathbf{y}$ $\widehat{\mathbf{a}} \to \mathbf{x} \circ \mathbf{y}$	$\widehat{\mathbf{h}} \to \mathbf{x}$ $\widehat{\mathbf{a}} \to \mathbf{y}$	$\widehat{\mathbf{h}} \to \mathbf{0}$ $\widehat{\mathbf{a}} \to \mathbf{0}$

Similarly, $\widehat{h}_i$ becomes proportional to $\mathbf{e}_i^T \mathbf{x} \left(\mathbf{x}^T \mathbf{z}_i + \mathbf{y}^T \mathbf{w}_i \right)$ and $\widehat{a}_i$ becomes proportional to $\mathbf{e}_i^T \mathbf{y} \left(\mathbf{x}^T \mathbf{w}_i + \mathbf{y}^T \mathbf{z}_i \right)$. We summarize in Table 1 what this implies in practice for the limiting behaviours of $\widehat{\mathbf{h}}$ and $\widehat{\mathbf{a}}$ in (5.3a) and (5.3b) according to the various possible choices of $\mathbf{w}_i$ and $\mathbf{z}_i$. The limits are reported ignoring global positive multiplicative constants. These do not affect the rankings induced by the centrality vectors, which are the objects of interest in network science. The symbol $\circ$ denotes the entrywise product.

From Table 1 it follows that selecting $\mathbf{z}_i = \mathbf{e}_i$ and $\mathbf{w}_i = \mathbf{0}$, i.e., a subgraph centrality type of measure, or $\mathbf{z}_i = \mathbf{0}$ and $\mathbf{w}_i = \mathbf{1}$, i.e., a total node communicability, would return the same ranking in the limit. Indeed, $\widehat{\mathbf{h}}$ will rank the nodes the same as $\mathbf{x}$ and $\widehat{\mathbf{a}}$ the same as $\mathbf{y}$. The same is true for all other choices of the vectors $\mathbf{z}_i, \mathbf{w}_i$, apart from when $\mathbf{w}_i = \mathbf{e}_i$ and the trivial choice $\mathbf{z}_i = \mathbf{w}_i = \mathbf{0}$.

7. Experiment with Synthetic Data. In this this section we describe some illustrative experiments on synthetic networks. This allows us to quantify the benefits of non-backtracking in a controlled setting. We consider the directed network represented by the adjacency matrix

$$A = \begin{bmatrix} 0 & B \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{n \times n}, \quad \text{where} \quad B = I + C = \begin{bmatrix} 1 & & & & 1 \\ 1 & 1 & & & \\ & 1 & \ddots & & \\ & & \ddots & \ddots & \\ & & & 1 & 1 \end{bmatrix} \in \mathbb{R}^{\frac{n}{2} \times \frac{n}{2}}, \tag{7.1}$$

with n even.

The nodes are ordered in such a way that $i \to \frac{n}{2} + i$ for all $i = 1, \ldots, \frac{n}{2}$, $i \to \frac{n}{2} + i - 1$ for all $i = 2, \ldots, \frac{n}{2}$ and $1 \to n$. Figure 1 illustrates the case where $n = 6$.

The nodes in this family of networks belong to two distinct groups: they are either sources or targets of edges. Following the notation of Section 3, we will denote the two groups as $S_1 = \{1, 2, \ldots, n/2\}$ and $S_2 = \{n/2 + 1, n/2 + 2, \ldots, n\}$. The network has been designed so that NBTAWs are forced to explore the network. For example, a NBTAW of length n must involve every node. For large n, eliminating backtracking should help to highlight pairs of nodes that are in the same bipartite group but far apart periodically (such as nodes 1 and $\lceil n/4 \rceil$). By contrast, a backtracking version

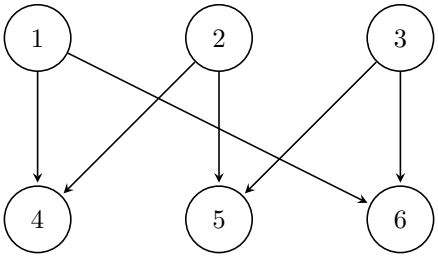

FIG. 1. *Example of the network with adjacency matrix A in (7.1) with $n = 6$.*

should give relatively high pairwise weights to nodes in the same bipartite group that are periodically close (such as nodes 1 and 2), since there are many more backtracking walks between them. So, overall, non-backtracking should give a more consistent weighting between all nodes in a bipartite group, especially when the downweighting parameter is large. For the same reason, the NBTAW approach should be less sensitive to perturbations in the structure.

The resolvent-based measures $G(A) + G(A^T)$ and $\widehat{G}(A) + \widehat{G}(A^T)$ are defined in terms of two positive downweighting parameters, α and t. In order for these measures to be well defined, we need $\alpha < 1/\rho(\mathcal{A})$ and $t < 1/\rho(Y)$, where Y is defined in (5.10). We next show that $\rho(\mathcal{A}) = 2$ and $\rho(Y) = 1$.

The matrix $\mathcal{A}$ is unitarily similar to the block-diagonal matrix $\operatorname{diag}(\Sigma, -\Sigma)$ where Σ is the diagonal matrix containing the singular values of A. It thus follows that $\rho(\mathcal{A}) = \sigma_1(A)$. Moreover, because of the structure of A, $\sigma_1(A) = \sigma_1(B)$ is the leading singular value of the circulant matrix B. We have

$$\sigma_1(B) = \sqrt{\lambda_1(BB^T)} = \sqrt{\lambda_1(2I + C + C^T)} = \sqrt{2 + \lambda_1(C + C^T)},$$

where we assume, here and in the following, that the eigenvalues and singular values are ordered in non-increasing modulus. The eigenvalues of the circulant matrix $C+C^T$ are $\lambda_j = 2\cos(\frac{2\pi j}{n/2})$ for $j = 1, 2, \ldots, n/2$. Hence, $\rho(C + C^T) = \lambda_1(C + C^T) = 2$ and $\rho(\mathcal{A}) = 2$.

Let us now consider the matrix Y, which is permutation-similar to

$$Y' := \left[\begin{array}{cccc|ccc} & & -I & B & & & \\ & & & I & & & \\ I & & & & & & \\ B^T & -I & & & & & \\ \hline & & & & & I & \\ & & & & & & I \\ & & & & I & & \\ & & & & & I & \end{array}\right].$$

Therefore, it follows that $\rho(Y) = \rho(Y') = \max\{1, \rho(Y'_{[1,1]})\}$, where $Y'_{[1,1]}$ denotes the leading $2n \times 2n$ block of Y'. This latter is permutation-similar to

$$\begin{bmatrix} A + A^T & -I \\ I & 0 \end{bmatrix},$$

which is the companion linearization [15] of the matrix polynomial $\operatorname{rev}(P(t))$ associated with the graph represented by $A + A^T$, where $P(t) = (1 - t^2)I - t(A + A^T)$.

This graph is simple and connected and all its nodes have degree exactly 2, so that the average degree of the nodes is 2. From [16, Lemma 6.2] it follows that $\mathrm{rev}(P(t))$, and thus $Y'_{[1,1]}$, has ν distinct finite eigenvalues $\lambda_j = \exp(-\frac{2\pi \hat{\imath} j}{\nu})$ for $j = 0, \ldots, \nu - 1$, where ν is the length of the unique cycle in the graph represented by $A + A^T$. Here, $\nu = n$ and thus $\rho(Y) = 1$.

The restrictions $\alpha < 1/2$ and $t < 1$ may also be understood intuitively by recalling that $G(A)$ is built from the resolvent function. For convergence of the power series, the factor α or t must control any increase in the alternating walk count from length k to $k + 1$. If we allow backtracking, then any alternating walk of length k spawns two alternating walks of length $k + 1$. Hence $\alpha < 1/2$ is necessary and sufficient to suppress this growth. Eliminating backtracking, only one of these two alternating walks of length $k + 1$ remains, so the constraint becomes $t < 1$.

Test 1. In this first test, we compare the performance of $F(A) + F(A^T)$ (resp., $G(A) + G(A^T)$) and $\widehat{F}(A) + \widehat{F}(A^T)$ (resp., $\widehat{G}(A) + \widehat{G}(A^T)$) in highlighting the sets S_1 and S_2 in a network with $n = 60$ nodes.

In Figure 2 we display heatmaps. As discussed in Section 3, large positive values in the diagonal blocks reveal intra-cluster relationships ($S_1 \to S_1$ and $S_2 \to S_2$), and large negative entries in the off-diagonal blocks reveal inter-cluster relationships ($S_1 \to S_2$ and $S_2 \to S_1$).

In the upper plots we see that both exponential-based measures, $F(A) + F(A^T)$ and $\widehat{F}(A) + \widehat{F}(A^T)$, show rapid decay away from the diagonal—because the measure emphasizes short walks, pairs of nodes in the same bipartite group that are periodically far apart are not highlighted. The color bar indicates that this effect is more pronounced for the standard walks.

For the lower plots in Figure 2, we see that the resolvent-based measures $F(A) + F(A^T)$ and $\widehat{F}(A) + \widehat{F}(A^T)$ with $\alpha = 0.99/2$ and $t = 0.99$ do a better job of revealing the structure, and in particular the non-backtracking version is able to highlight all types of interaction.

To give further detail, in Figure 3 we show heatmaps for different values of $\alpha = p/2$ and $t = p$, where $p = 0.25, 0.5, 0.8, 0.99$. From these pictures it can be seen that the non-backtracking measure displays a smoother off-diagonal decay in the entries, especially for larger values of the downweighting parameter.

Test 2. We now quantify the resilience of these methods to noise, in the form of spurious edges that impair the bipartite structure. More precisely, we successively added an extra directed edge to the network up to a limit of 60 edges. Each new edge was chosen uniformly at random, with the condition that repeated edges and loops are not allowed. After each edge had been added, we computed the symmetric matrices $G(A) + G(A^T)$ and $\widehat{G}(A) + \widehat{G}(A^T)$ for the new adjacency matrix A. To break the network into two groups, we used a standard spectral clustering approach [18, 33]—we computed the eigenvector $\mathbf{v}^{[1]}$ associated with the dominant eigenvalue λ_1 and assigned node i and node j to the same group if $v_i^{[1]}$ and $v_j^{[1]}$ shared the same sign. This splits the nodes into two groups. In order to judge the algorithms, we regarded the group with the most S_1 nodes as our approximation to $S_1 = \{1, 2, \ldots, n/2 = 30\}$ and the other group as our approximation to $S_2 = \{n/2 + 1, n/2 + 2, \ldots, n = 60\}$. Note that the algorithm was not forced to place 30 nodes in each group; we were not hard-wiring the group size into the tests.

We assessed performance with the *F1 score*, which is the harmonic average of

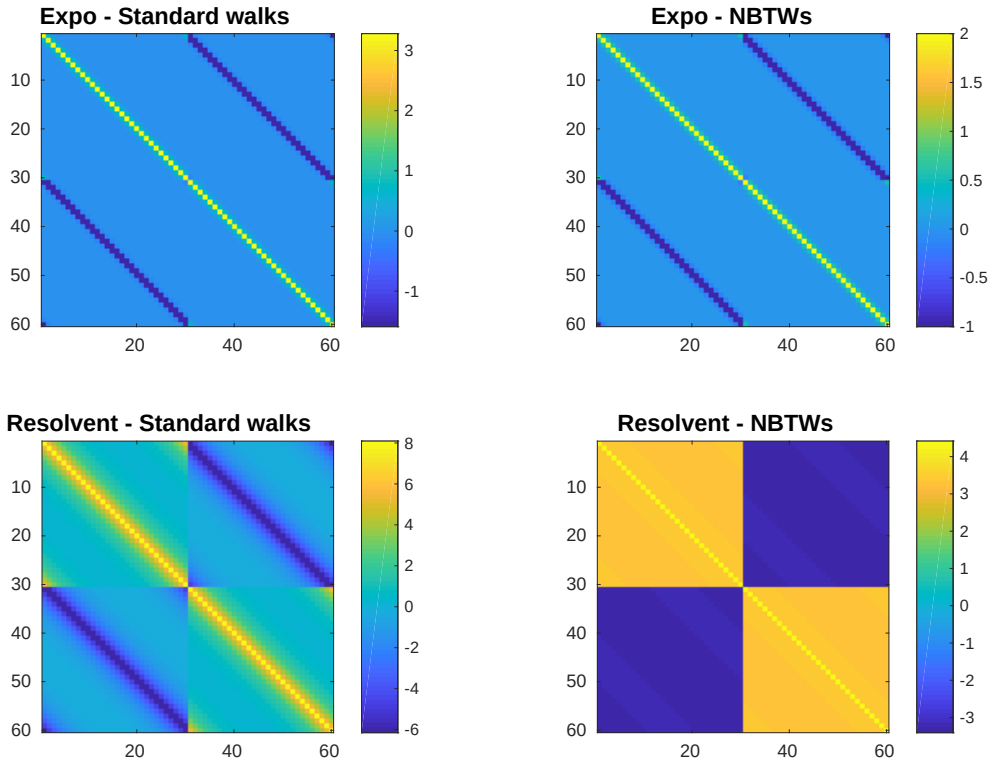

FIG. 2. *Upper: heatmaps of* $F(A) + F(A^T)$ *and* $\widehat{F}(A) + \widehat{F}(A^T)$. *Lower: heatmaps of* $G(A) + G(A^T)$ *(*$\alpha = 0.99/2$*) and* $\widehat{G}(A) + \widehat{G}(A^T)$ *(*$t = 0.99$*).*

precision and recall:

$$\mathrm{F1} = 2\,\frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}.$$

The F1 score ranges between 0 and 1, with 1 representing perfect precision and recall. Recall that *precision* is the ratio of the number of relevant items to the number of those selected by the method, so

$$\text{precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}},$$

whilst *recall* (or *sensitivity*) is the ratio of the number of relevant items selected to the overall number of relevant items, and thus

$$\text{recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}.$$

In more detail, we display the F1 score relative to the identification of the nodes in the set S_1. In this case, precision is the ratio between the number of positive (or negative) entries found in the top 30 entries of the eigenvector considered, i.e., the correctly identify nodes, and the total number of its entries with that sign. The recall, on the other hand, is the ratio of the number of correctly identified nodes to the size of S_1. (Clearly, good performance with respect to S_1 corresponds to good performance with respect S_2.)

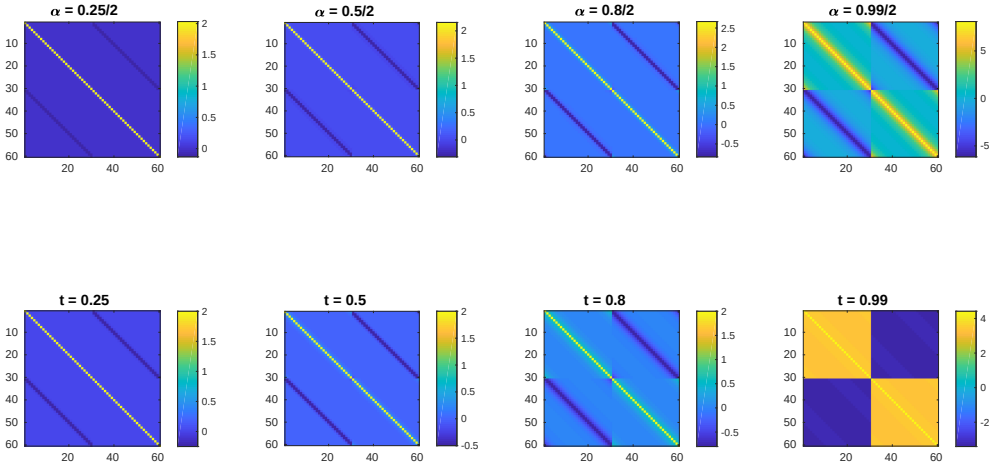

FIG. 3. *Heatmaps of* $G(A) + G(A^T)$ *(upper) and* $\widehat{G}(A) + \widehat{G}(A^T)$ *(lower) for different values of* α *and* t*.*

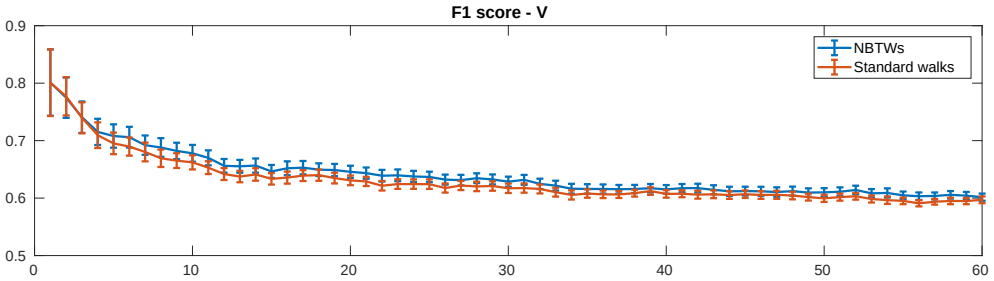

FIG. 4. *Evolution of the F1 score of the partition induced by the eigenvectors associated with the leading eigenvalues of* $G(A) + G(A^T)$ *and* $\widehat{G}(A) + \widehat{G}(A^T)$*. Here* α *and* t *are chosen to be* 0.25 *of their upper limits.*

We ran this test 250 times and averaged the results. In Figures 4–6 we display the evolution of the average F1 score with the number of updates performed, with standard errors appended.

The parameters α and t used in the computations were $0.25, 0.5, 0.99$ of the inverse of the spectral radius of $\mathcal{A}$ or Y, respectively, and were updated after each rank-1 modification of the adjacency matrix. The plots show that the non-backtracking version of the measure induced by the matrix resolvent is more resilient to the presence of noise, as it returns on average higher values for the F1 score in all cases.

8. Experiment with Worm Brain Data. We now test the performance of the measure $\widehat{F}(A)+\widehat{F}(A^T)$ on a real network. We use a local subnetwork of 131 nodes from the nematode (roundworm) *Caenorhabditis elegans*. Here, nodes represent neurons and edges reflect physical connections. This network was analyzed in [12] using $F(A)+F(A^T)$, where it was shown that discovering directed bipartite substructure revealed insights into the command structure within the organism's nervous system. The authors identified two sets S_1 and S_2 of 16 nodes each that constitute an approximate directed bipartite subgraph in the network: the density of the submatrix $S_1 \to S_2$ being at least five times larger that of $S_i \to S_i$, for $i = 1, 2$ and $S_2 \to S_1$. A heatmap of the reordered $F(A) + F(A^T)$ is displayed in Figure 7 and the hot zones in

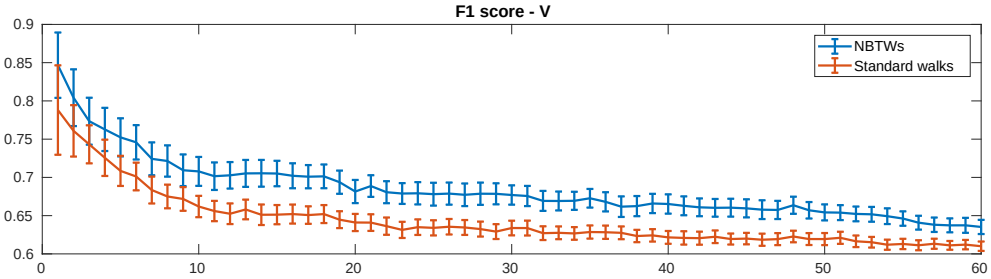

FIG. 5. *Evolution of the F1 score of the partition induced by the eigenvectors associated with the leading eigenvalues of* $G(A)+G(A^T)$ *and* $\widehat{G}(A)+\widehat{G}(A^T)$. *Here* α *and* t *are chosen to be* 0.5 *of their upper limits.*

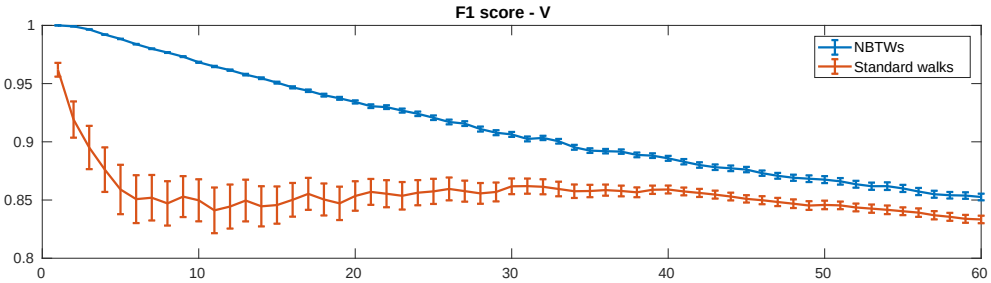

FIG. 6. *Evolution of the F1 score of the partition induced by the eigenvectors associated with the leading eigenvalues of* $G(A)+G(A^T)$ *and* $\widehat{G}(A)+\widehat{G}(A^T)$. *Here* α *and* t *are chosen to be* 0.99 *of their upper limits.*

the upper left and lower right corners correspond to the indices associated with the nodes in S_1 and S_2, respectively. In Figure 8 we display a heatmap of the reordered $\widehat{F}(A)+\widehat{F}(A^T)$. The two hot regions represent two sets $\widehat{S}_1$ and $\widehat{S}_2$ containing 24 and 16 nodes, respectively. The density of the submatrix $\widehat{S}_1 \to \widehat{S}_2$ is 4, 16, and 32 times the densities of the submatrices $\widehat{S}_1 \to \widehat{S}_1$, $\widehat{S}_2 \to \widehat{S}_2$, and $\widehat{S}_2 \to \widehat{S}_1$. Figure 9 displays surface plots of $F(A)+F(A^T)$ (left) and its non-backtracking analogue (right). As observed above, the largest entries of the non-backtracking version $\widehat{F}(A)+\widehat{F}(A^T)$ are smaller in modulus that those of $F(A)+F(A^T)$ but they drop off less rapidly away from the diagonal, giving more clearly defined clusters.

In summary, the new non-backtracking version of the algorithm has discovered larger and denser directed bipartite substructure than the original method.

9. Summary. Our aim in this work was to combine two concepts that have proved useful in the development of walk-based algorithms for networks: *non-backtracking* allows the network to be explored more thoroughly and *alternating* reveals bipartite or hub/authority structures. We developed the required combinatoric theory for counting non-backtracking alternating walks and showed that convenient expressions can be derived for the associated power series. This enabled efficient algorithms to be devised—the backtracking constraint essentially imposes no extra cost. We also derived the parameter-free limit of the resolvent-based hub/authority measure, giving an analogue of the classical spectral measures.

REFERENCES

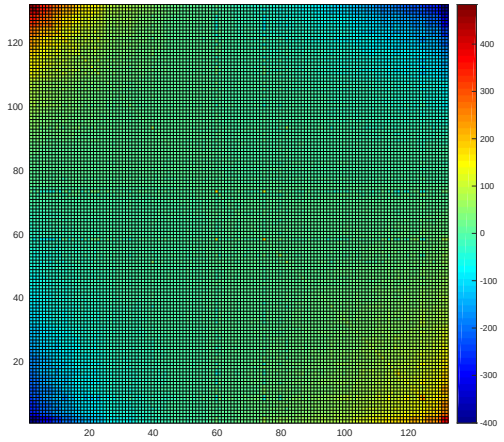

FIG. 7. *Heatmap of the reordered version of $F(A) + F(A^T)$, reordered using the entries of the eigenvector associated with its largest eigenvalue.*

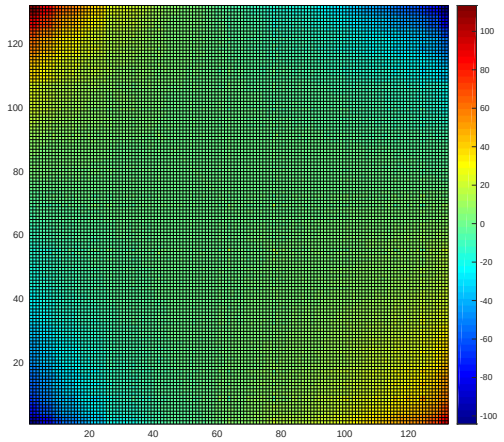

FIG. 8. *Heatmap of the reordered version of $\widehat{F}(A) + \widehat{F}(A^T)$, reordered using the entries of the eigenvector associated with its largest eigenvalue.*

[1] N. Alon, I. Benjamini, E. Lubetzky, and S. Sodin. Non-backtracking random walks mix faster. *Communications in Contemporary Mathematics*, 09:585–603, 2007.
[2] O. Angel, J. Friedman, and S. Hoory. The non-backtracking spectrum of the universal cover of a graph. *Transactions of the American Mathematical Society*, 326:4287–4318, 2015.
[3] F. Arrigo and M. Benzi. Edge modification criteria for enhancing the communicability of digraphs. *SIAM Journal on Matrix Analysis and Applications*, 37(1):443–468, 2016.
[4] F. Arrigo, M. Benzi, and C. Fenu. Computation of generalized matrix functions. *SIAM Journal on Matrix Analysis and Applications*, 37(3):836–860, 2016.
[5] F. Arrigo, P. Grindrod, D. J. Higham, and V. Noferini. On the exponential generating function for non-backtracking walks. (submitted), 2017.
[6] F. Arrigo, P. Grindrod, D. J. Higham, and V. Noferini. Non-backtracking walk centrality for directed networks. *Journal of Complex Networks*, 6(1):54–78, 2018.
[7] M. Benzi, E. Estrada, and C. Klymko. Ranking hubs and authorities using matrix functions. *Linear Algebra and its Applications*, 438(5):2447–2474, 2013.
[8] M. Benzi and C. Klymko. Total communicability as a centrality measure. *Journal of Complex Networks*, 1(2):124–149, 2013.
[9] M. Benzi and C. Klymko. On the limiting behavior of parameter-dependent network centrality measures. *SIAM J. Matrix Anal. Appl.*, 36:686–706, 2015.

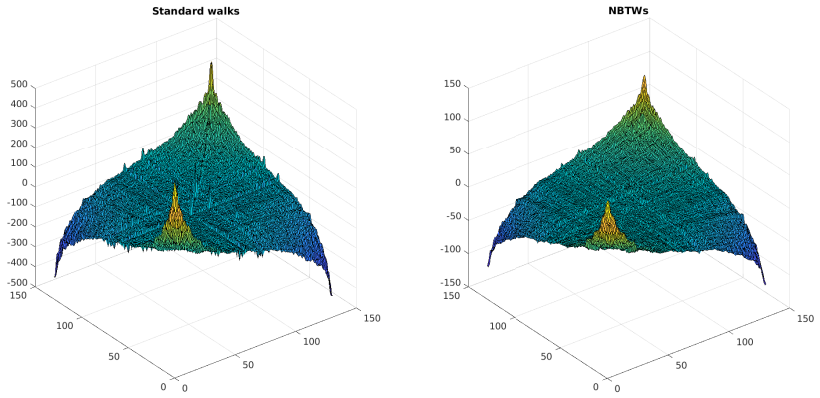

FIG. 9. *Surface plot of the reordered version of* $F(A) + F(A^T)$ *(left) and of* $\widehat{F}(A) + \widehat{F}(A^T)$ *(right)*

[10] R. Bowen and O. E. Lanford. Zeta functions of restrictions of the shift transformation. In S.-S. Chern and S. Smale, editors, *Global Analysis: Proceedings of the Symposium in Pure Mathematics of the Americal Mathematical Society, University of California, Berkely, 1968*, pages 43–49. American Mathematical Society, 1970.

[11] R. A. Brualdi, F. Harary, and Z. Miller. Bigraphs versus digraphs via matrices. *Journal of Graph Theory*, 4(1):51–73, 1980.

[12] J. J. Crofts, E. Estrada, D. J. Higham, and A. Taylor. Mapping directed networks. *Electron. Trans. Numer. Anal*, 37:337–350, 2010.

[13] F. De Terán, F. M. Dopico, and P. Van Dooren. Matrix polynomials with completely prescribed eigenstructure. *SIAM Jorunal on Matrix Analysis and Applications*, 36:302–328, 2015.

[14] E. Estrada and D. J. Higham. Network properties revealed through matrix functions. *SIAM Review*, 52:696–671, 2010.

[15] I. Gohberg, P. Lancaster, and L. Rodman. *Matrix Polynomials*. SIAM, Philadelphia, PA, 2009.

[16] P. Grindrod, D. J. Higham, and V. Noferini. The deformed graph Laplacian and its applications to network centrality analysis. *SIAM Journal on Matrix Analysis and Applications*, 39(1), 2018.

[17] J. B. Hawkins and A. Ben-Israel. On generalized matrix functions. *Linear and Multilinear Algebra*, 1(2):163–171, 1973.

[18] D. J. Higham, G. Kalna, and J. K. Vass. Spectral analysis of two-signed microarray expression data. *Mathematical Medicine and Biology*, 24(2):131–148, 2007.

[19] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, 1985.

[20] M. D. Horton. Ihara zeta functions on digraphs. *Linear Algebra and its Applications*, 425:130–142, 2007.

[21] M. D. Horton, H. M. Stark, and A. A. Terras. What are zeta functions of graphs and what are they good for? In G. Berkolaiko, R. Carlson, S. A. Fulling, and P. Kuchment, editors, *Quantum graphs and their applications*, volume 415 of *Contemp. Math.*, pages 173–190. 2006.

[22] T. Kawamoto. Localized eigenvectors of the non-backtracking matrix. *Journal of Statistical Mechanics: Theory and Experiment*, 2016:023404, 2016.

[23] J. M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, 46(5):604–632, 1999.

[24] F. Krzakala, C. Moore, E. Mossel, J. Neeman, A. Sly, L. Zdeborová, and P. Zhang. Spectral redemption: clustering sparse networks. *Proceedings of the National Academy of Sciences*, 110:20935–20940, 2013.

[25] T. Martin, X. Zhang, and M. E. J. Newman. Localization and centrality in networks. *Phys. Rev. E*, 90:052808, 2014.

[26] Y. Nakatsukasa, V. Noferini, and A. Townsend. Vector spaces of linearizations of matrix

polynomials: a bivariate polynomial approach. *SIAM J. Matrix Anal. Appl.*, 38(1):1–29, 2016.

[27] R. Pastor-Satorras and C. Castellano. Distinct types of eigenvector localization in networks. *Scientific Reports*, 6:18847, 2016.

[28] Y. Saad. Analysis of some Krylov subspace approximations to the matrix exponential operator. *SIAM Journal on Numerical Analysis*, 29(1):209–228, 1992.

[29] A. Saade, F. Krzakala, and L. Zdeborová. Spectral clustering of graphs with the Bethe Hessian. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 406–414. 2014.

[30] H. Stark and A. Terras. Zeta functions of finite graphs and coverings. *Advances in Mathematics*, 121(1):124–165, 1996.

[31] A. Tarfulea and R. Perlis. An Ihara formula for partially directed graphs. *Linear Algebra and its Applications*, 431:73–85, 2009.

[32] H. Taubig. *Matrix Inequalities for Iterative Systems*. CRC Press, 2017.

[33] A. Taylor, J. K. Vass, and D. J. Higham. Discovering bipartite substructure in directed networks. *LMS Journal of Computation and Mathematics*, 14:72–86, 2011.

[34] A. Terras. *Harmonic Analysis on Symmetric Spaces — Euclidean Space, the Sphere, and the Poincaré Upper Half-Plane*. Springer, New York, 2nd edition, 2013.

[35] Y. Watanabe and K. Fukumizu. Graph zeta function in the Bethe free energy and loopy belief propagation. In Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams, and A. Culotta, editors, *Advances in Neural Information Processing Systems 22*, pages 2017–2025. 2009.