

*On the exponential generating function for
non-backtracking walks*

Arrigo, Francesca and Grindrod, Peter and Higham,
Desmond J. and Noferini, Vanni

2017

MIMS EPrint: **2017.34**

Manchester Institute for Mathematical Sciences
School of Mathematics

The University of Manchester

Reports available from: <http://eprints.maths.manchester.ac.uk/>

And by contacting: The MIMS Secretary
School of Mathematics
The University of Manchester
Manchester, M13 9PL, UK

ISSN 1749-9097

ON THE EXPONENTIAL GENERATING FUNCTION FOR NON-BACKTRACKING WALKS

FRANCESCA ARRIGO[†] PETER GRINDROD[‡] DESMOND J. HIGHAM[§] AND VANNI
NOFERINI[¶]

Abstract. We derive an explicit formula for the exponential generating function associated with non-backtracking walks around a graph. We study both undirected and directed graphs. Our results allow us to derive computable expressions for non-backtracking versions of network centrality measures based on the matrix exponential. We find that eliminating backtracking walks in this context does not significantly increase the computational expense. We show how the new measures may be interpreted in terms of standard exponential centrality computation on a certain multilayer network. Insights from this block matrix interpretation also allow us to characterize centrality measures arising from general matrix functions. Rigorous analysis on the star graph illustrates the effect of non-backtracking and shows that unwanted localization effects can be eliminated when we restrict to non-backtracking walks. We also investigate the localization issue on synthetic networks.

Key words. generating function, inverse participation ratio, Katz, localization, loops, multi-layer, graph spectrum, backtracking walks.

AMS subject classifications. 05C90, 05C50, 15A15, 65F50

1. Introduction.

1.1. Motivation. The concept of a walk on a graph is very natural—on arriving at a node, the walker may continue by traversing any edge pointing out of that node. If, at any stage, the edge along which the walker continues is the reverse of the edge on which they arrived, then the walk is said to be *backtracking*. Non-backtracking walks have been analyzed in a number of fields. They play a key role in the study of zeta functions on graphs [24], with applications in spectral graph theory [4, 23], number theory [44], discrete mathematics [12, 41], quantum chaos [39], random matrix theory [40], stochastic analysis [3], applied linear algebra [42] and computer science [38, 45]. Within network science, constraining walks to be non-backtracking has been shown to offer benefits in community detection [26, 28], centrality measurement [6, 20, 29, 36], network comparison [30] and in the study of related issues concerning optimal percolation [31, 32].

A natural task across all these fields is to count the number of distinct non-backtracking walks of a given length between pairs of nodes, and to form a compact expression for the associated generating function. For network centrality, in the case of standard walk counts it has been argued that an exponential-style power series gives a useful alternative to the standard resolvent-style version [15]. For example, based on an analogy with a physical oscillator, it may be argued that moving from the resolvent to the exponential takes us from classical to quantum physics [16]. Further, there are effective and reliable tools for computing the action of the matrix exponential [2, 21, 22]. This provides the initial motivation for our article, where we study the exponential generating function associated with non-backtracking walks. We also note that more general matrix functions have been proposed in this walk-counting

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

[‡]Mathematical Institute, University of Oxford, Oxford, UK (peter.grindrod@maths.ox.ac.uk)

[§]Department of Mathematics and Statistics, University of Strathclyde, Glasgow, UK (d.j.higham@strath.ac.uk)

[¶]Department of Mathematical Sciences, University of Essex, Colchester, UK (vnofer@essex.ac.uk)

context [17]. Hence, we then extend the analysis to cover generating functions based on arbitrary power series.

1.2. Background. We let $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ denote the adjacency matrix for an unweighted graph with n nodes; that is, $a_{ij} = 1$ if there is an edge from node i to node j , and $a_{ij} = 0$ otherwise. We will assume the graph to be loopless, so that $a_{ii} = 0$ for all i , and to have no multiple edges. If we further assume the graph represented by A to be *undirected*, i.e., if $A = A^T$, then the graph will be said to be *simple*. We use D to denote the diagonal matrix whose diagonal entries are $D_{ii} = (A^2)_{ii}$. So D_{ii} counts the number of reciprocal neighbours of node i , that is, the number of nodes j such that $a_{ij} = a_{ji} = 1$. If $A = A^T$, so that all edges are reciprocated, then D_{ii} reduces to the vector of *degrees* and any node i for which $D_{ii} = 1$ will be called a *leaf*.

If $A \neq A^T$ the network is said to be *directed* and we will denote by $S = (s_{ij}) \in \mathbb{R}^{n \times n}$ the matrix associated with the graph obtained by removing all edges that are not reciprocated, so that $s_{ij} = a_{ij}a_{ji}$ for all $i, j = 1, 2, \dots, n$.

A *walk* of length r from node i to node j is characterized by a sequence of edges $i = i_1 \mapsto i_2, i_2 \mapsto i_3, \dots, i_r \mapsto i_{r+1} = j$. Note that edges and nodes may be revisited in the course of a walk. It follows from the definition of matrix multiplication that A^r has (i, j) th element that counts the number of distinct walks of length r from i to j ; see, for example, [13, Theorem 2.2.1]. Given a parameter t , the *ordinary generating function* [46] associated with this sequence of walk counts may be represented by a resolvent function:

$$\sum_{r=0}^{\infty} t^r A^r = (I - tA)^{-1},$$

where I is the identity matrix. If t is interpreted as a formal algebraic variable, this expression holds within the algebra of formal power series. Analytically, if t is interpreted as a real (or complex) valued parameter, it is valid for $|t| < 1/\rho(A)$, where $\rho(\cdot)$ denotes the spectral radius. Similarly, the associated *exponential generating function* [46] takes the form of a matrix exponential:

$$\sum_{r=0}^{\infty} \frac{t^r A^r}{r!} = \exp(tA),$$

analytically valid for all $t \in \mathbb{R}$ (or in \mathbb{C}).

Power series expansions that combine walk counts of different lengths have become widely used in the field of network science [17]. Here, the idea is to assign a weight to each node that summarizes its “importance” in the network. If we interpret importance as the ability of a node to initiate messages along the edges in the graph, then we see that $(I - tA)^{-1}\mathbf{1}$, for $0 < t < 1/\rho(A)$, and $\exp(tA)\mathbf{1}$, for $t > 0$, are candidates for measuring and comparing nodes, where $\mathbf{1} \in \mathbb{R}^n$ is the vector of ones. Here, the i th component arises by summing the number of distinct walks from node i to all nodes in the graph, with walks of length r discounted by t^r in the resolvent case and by $t^r/r!$ in the exponential case. The resolvent version dates back to the work of Katz [25] and the alternative based on the exponential function, called the *total (node) communicability*, was introduced and studied in [9]. We also note that both versions are related to eigenvector centrality [10, 11], which uses the Perron-Frobenius eigenvector of A when the underlying graph is strongly connected [8].

In recent work [29], Martin et al. argued that traditional centrality measures may suffer from unwanted *localization* effects when the network under study has a

scale-free degree distribution (see, e.g., [7, 14]). In this type of network, there are a few nodes with a high number of neighbours, that are thus assigned most of the centrality, and many nodes with a very small number of connections, that hence share the remainder of the mass of the centrality vector. This unbalanced partitioning makes it difficult to distinguish between the importance of the nodes with few connections. The authors therefore proposed a variation of eigenvector centrality motivated by the concept of non-backtracking walks. This idea was pursued in [6, 20] in the context of resolvent-based centrality. Our aim here is to develop and study the analogous exponential version of non-backtracking walk centrality, and also to consider general matrix functions.

Formally, we say that a walk is *backtracking* if it contains at least one pair of successive edges of the form $i \mapsto j, j \mapsto i$; that is, having left some node i it immediately returns to that node. We say the walk is *non-backtracking* otherwise. We use the abbreviation NBTW to mean non-backtracking walk. We will denote by $p_r(A) \in \mathbb{R}^{n \times n}$ the matrix whose (i, j) th entry counts the number of NBTWs of length r from node i to node j in the graph represented by the matrix A . Our first task is therefore to obtain useful expressions for the corresponding exponential generating function

$$(1.1) \quad F(t) := \sum_{r=0}^{\infty} \frac{t^r p_r(A)}{r!}$$

and centrality measure

$$(1.2) \quad \mathbf{b} := F(t)\mathbf{1}.$$

We do this in Sections 2 and 3 for the case of undirected and directed graphs, respectively. Although, of course, an undirected graph is a special case of a directed graph, we treat the two cases separately for two reasons. First, undirected graphs are common in network science, and hence it is useful to have results tailored to this case. Second, in our context it seems more straightforward to study undirected graphs directly than to simplify the more general expression for directed graphs. In Section 4, we briefly describe how to compute the proposed centrality measures and discuss computational costs. In Section 5, we interpret the new expressions from the viewpoint of traversals around multilayer networks with possibly negative interlayer weights. This block matrix framework also allows us to derive expressions for the case where the exponential is replaced by a general matrix function. In Section 6 we give a concrete example of the benefit of non-backtracking, and further illustrations on realistic model networks appear in Section 7. We end with a short summary in Section 8.

The key analytical steps used here are to derive and solve an ODE—this is a widely used approach in the study of exponential generating functions [46]. To this end, we note from (1.1) that

$$(1.3) \quad F'(t) = \sum_{r=0}^{\infty} \frac{t^r p_{r+1}(A)}{r!} \quad \text{and} \quad F''(t) = \sum_{r=0}^{\infty} \frac{t^r p_{r+2}(A)}{r!}.$$

We also require a lemma on matrix-valued ODEs. The following result arises from columnwise application of a standard result on vector-valued linear systems; see, for example, [21, Section 2.1]. The statement of the result makes use of the shifted exponential function $\psi_1(z) = (e^z - 1)/z$, which may be defined formally via the power

series expansion

$$(1.4) \quad \psi_1(z) = \sum_{r=0}^{\infty} \frac{z^r}{(r+1)!}.$$

LEMMA 1.1. *Let $C \in \mathbb{R}^{m \times m}$ and $P, X_0 \in \mathbb{R}^{m \times \ell}$. Then*

$$(1.5) \quad X(t) = t\psi_1(tC)(P + CX_0) + X_0$$

is the unique solution to the linear, constant coefficient, inhomogeneous, autonomous, initial value ODE system

$$X'(t) = CX(t) + P, \quad X(0) = X_0.$$

Moreover, in the homogeneous case $P = 0$, the expression (1.5) simplifies to

$$(1.6) \quad X(t) = \exp(tC)X_0.$$

REMARK 1.2. *We note that Lemma 1.1 does not require the Jacobian matrix C to be nonsingular.*

2. Exponential generating function for undirected graphs. We assume throughout this section that the graph is undirected. We also assume for convenience that the graph is connected—in the disconnected case, each component may be studied separately. We note that each node is therefore assumed to have at least one neighbour.

As we mentioned in Section 1.2, the matrix A^r counts walks of length r . Hence the trivial relationship $A^{r+1} = AA^r$ gives a recurrence between walk counts for length r and $r + 1$. In the case of NBTWs on undirected graphs, the following two term recurrence relates walk counts of successive lengths. We have $p_1(A) = A$, $p_2(A) = A^2 - D$, and

$$(2.1) \quad p_{r+2}(A) = Ap_{r+1}(A) - (D - I)p_r(A), \quad \text{for } r \geq 1.$$

This recursion was proved in [12] for general directed graphs. For the undirected case considered here, the recurrence was re-derived independently in [41].

We now give the main statement of this section: an explicit formula for the generating function (1.1) and the corresponding centrality measure (1.2).

THEOREM 2.1. *For a simple graph with associated adjacency matrix A , the exponential generating function $F(t)$ in (1.1) and total communicability vector \mathbf{b} in (1.2) satisfy*

$$F(t) = \begin{bmatrix} I & 0 \end{bmatrix} t\psi_1(tY) \begin{bmatrix} A \\ A^2 - D \end{bmatrix} + I,$$

$$\mathbf{b} = \begin{bmatrix} I & 0 \end{bmatrix} t\psi_1(tY) \begin{bmatrix} A\mathbf{1} \\ (A^2 - D)\mathbf{1} \end{bmatrix} + \mathbf{1},$$

where

$$(2.2) \quad Y := \begin{bmatrix} 0 & I \\ I - D & A \end{bmatrix} \in \mathbb{R}^{2n \times 2n}.$$

In the case where the graph has no leaves, these expressions may be written as

$$\begin{aligned} F(t) &= \begin{bmatrix} I & 0 \end{bmatrix} \exp(tY) \begin{bmatrix} I + (D - I)^{-1} \\ A \end{bmatrix} - (D - I)^{-1}, \\ \mathbf{b} &= \begin{bmatrix} I & 0 \end{bmatrix} \exp(tY) \begin{bmatrix} \mathbf{1} + (D - I)^{-1}\mathbf{1} \\ A\mathbf{1} \end{bmatrix} - (D - I)^{-1}\mathbf{1}. \end{aligned}$$

Proof. To obtain the exponential generating function $F(t)$ in (1.1), we multiply by $t^r/r!$ in (2.1) and then sum from $r = 1$ to ∞ , to obtain

$$\sum_{r=1}^{\infty} \frac{t^r p_{r+2}(A)}{r!} = A \sum_{r=1}^{\infty} \frac{t^r p_{r+1}(A)}{r!} - (D - I) \sum_{r=1}^{\infty} \frac{t^r p_r(A)}{r!}.$$

From (1.3), this shows

$$F''(t) - p_2(A) = A(F'(t) - p_1(A)) - (D - I)(F(t) - p_0(A)),$$

which simplifies to

$$(2.3) \quad F''(t) - AF'(t) + (D - I)F(t) = -I.$$

To convert to a first order system, we introduce

$$G(t) = \begin{bmatrix} G_1(t) \\ G_2(t) \end{bmatrix} := \begin{bmatrix} F(t) \\ F'(t) \end{bmatrix} \in \mathbb{R}^{2n \times n},$$

so that (2.3) becomes

$$G'(t) = YG(t) + \begin{bmatrix} 0 \\ -I \end{bmatrix}, \quad \text{where } G(0) = \begin{bmatrix} I \\ A \end{bmatrix},$$

where Y is defined in (2.2). It then follows from Lemma 1.1 that

$$(2.4) \quad G(t) = t\psi_1(tY) \left(\begin{bmatrix} 0 \\ -I \end{bmatrix} + Y \begin{bmatrix} I \\ A \end{bmatrix} \right) + \begin{bmatrix} I \\ A \end{bmatrix}.$$

Finally, we note that the Jacobian matrix Y in (2.2) is invertible if and only if $(D - I)^{-1}$ exists, that is, if the graph does not contain leaves, i.e., nodes of degree one. In this case

$$Y^{-1} = \begin{bmatrix} (D - I)^{-1}A & -(D - I)^{-1} \\ I & 0 \end{bmatrix}$$

and (2.4) simplifies to

$$(2.5) \quad G(t) = \exp(tY) \begin{bmatrix} I + (D - I)^{-1} \\ A \end{bmatrix} - \begin{bmatrix} (D - I)^{-1} \\ 0 \end{bmatrix}.$$

Noting that $F(t) = \begin{bmatrix} I & 0 \end{bmatrix} G(t)$, we see that (2.4) and (2.5) yield the statement.

□

We discuss these results further in Sections 4 and 5.

3. Exponential generating function for directed graphs. For directed graphs the matrices $p_r(A)$ that count NBTWs satisfy a three term recurrence. This was shown in [12], with a more linear algebra oriented derivation given in [42]. With $p_0(A) = I$, $p_1(A) = A$, $p_2(A) = A^2 - D$, the recurrence takes the form

$$(3.1) \quad p_{r+3}(A) = Ap_{r+2}(A) - (D - I)p_{r+1}(A) - (A - S)p_r(A), \quad \text{for } r \geq 0.$$

Recall that $S \in \mathbb{R}^{n \times n}$ is the adjacency matrix of the graph obtained by only considering reciprocated links in the original network; so $s_{ij} = a_{ij}a_{ji}$. We emphasize that unlike the undirected case (2.1), this recurrence is valid from $r = 0$.

Here is the analogue for directed graphs of Theorem 2.1.

THEOREM 3.1. *Let A be the adjacency matrix of an unweighted directed graph with no self-loops nor multiple edges. The exponential generating function $F(t)$ in (1.1) and total communicability vector \mathbf{b} in (1.2) satisfy*

$$F(t) = \begin{bmatrix} I & 0 & 0 \end{bmatrix} \exp(tZ) \begin{bmatrix} I \\ A \\ A^2 - D \end{bmatrix},$$

$$\mathbf{b} = \begin{bmatrix} I & 0 & 0 \end{bmatrix} \exp(tZ) \begin{bmatrix} \mathbf{1} \\ A\mathbf{1} \\ (A^2 - D)\mathbf{1} \end{bmatrix},$$

where

$$(3.2) \quad Z := \begin{bmatrix} 0 & I & 0 \\ 0 & 0 & I \\ -(A - S) & -(D - I) & A \end{bmatrix} \in \mathbb{R}^{3n \times 3n}.$$

Proof. Multiplying (3.1) by $t^r/r!$ and summing from $r = 0$ to ∞ we arrive at

$$(3.3) \quad F'''(t) = AF''(t) - (D - I)F'(t) - (A - S)F(t).$$

We note that, in contrast to (2.3), this ODE does not have an inhomogeneous term. To convert to a first order system, introduce

$$H(t) = \begin{bmatrix} H_1(t) \\ H_2(t) \\ H_3(t) \end{bmatrix} = \begin{bmatrix} F(t) \\ F'(t) \\ F''(t) \end{bmatrix},$$

so that

$$H'(t) = ZH(t), \quad \text{where } H(0) = \begin{bmatrix} I \\ A \\ A^2 - D \end{bmatrix},$$

where Z is defined in (3.2).

Using (1.6) in Lemma 1.1 and noting that $F(t) = \begin{bmatrix} I & 0 & 0 \end{bmatrix} H(t)$, we arrive at the statement.

□

It is of interest to note that the removal of backtracking walks may be interpreted as the introduction of damping terms $-(D - I)F'(t)$ and $-(A - S)F(t)$ in (3.3). Further, Theorem 3.1 shows that, in general, as t increases the total communicability vector \mathbf{b} aligns with the first n components of the dominant eigenvector of Z .

4. Computing the centrality vectors. In this section we consider computational issues. First, we note that the standard expression for the total (node) communicability, $e^{tA}\mathbf{1}$, requires the action of the exponential of a sparse n by n matrix on a vector in \mathbb{R}^n . It is reasonable to assume that a suitable approximation may be obtained iteratively, using a small number of sparse matrix-vector multiplications [2]. Assuming that A has $O(n)$ nonzeros (edges), then each sparse matrix-vector multiplication will cost $O(n)$. Hence, using k to denote the number of iterations, the overall cost will be $O(nk)$. In practice, it is reasonable to regard k as finite and independent from n , so that this expression becomes $O(n)$.

In Theorem 2.1, we see that for an undirected graph with no leaves, the cost of computing the non-backtracking total communicability vector \mathbf{b} is dominated by the action of e^{tY} on a vector in \mathbb{R}^{2n} , where the $2n$ by $2n$ sparse matrix Y is defined in (2.2). We note that Y has only $2n$ more nonzeros than the underlying adjacency matrix A , and hence, under the assumptions above, if the same number of iterations are required, then the same $O(n)$ cost applies.

In the directed case, the expression for \mathbf{b} in Theorem 3.1 requires the action of e^{tZ} on a vector in \mathbb{R}^{3n} , where the $3n$ by $3n$ sparse matrix Z is defined in (3.2). Since Z also has $O(n)$ nonzeros, we again recover the $O(n)$ cost under our assumptions.

It remains to consider the case of an undirected graph with leaves. In Theorem 2.1, the expression for \mathbf{b} involves the shifted exponential, ψ_1 , from (1.4). We may proceed by reducing the key computation to the approximation of the action of a matrix exponential in a slightly higher dimension. Let

$$\widehat{Y} := \begin{bmatrix} tY & \mathbf{v} \\ \mathbf{0}^T & 0 \end{bmatrix} \in \mathbb{R}^{(2n+1) \times (2n+1)},$$

where $\mathbf{v} \in \mathbb{R}^{2n}$ and $\mathbf{0} \in \mathbb{R}^{2n}$ is a vector of all zeros. It was shown in [37, Section 2.3] that

$$\exp(\widehat{Y}) = \begin{bmatrix} e^{tY} & \psi_1(tY)\mathbf{v} \\ \mathbf{0}^T & 1 \end{bmatrix}.$$

So, if we let

$$\mathbf{v} = \begin{bmatrix} A\mathbf{1} \\ (A^2 - D)\mathbf{1} \end{bmatrix}$$

then \mathbf{b} in Theorem 2.1 may be computed by first approximating the vector $\exp(\widehat{Y})\mathbf{e}_{2n+1}$, where \mathbf{e}_{2n+1} is the $(2n+1)$ th vector of the standard basis of \mathbb{R}^{2n+1} . We then extract the top n entries of the resulting vector, scale by t and shift by $\mathbf{1}$. This again will have an $O(n)$ cost under our assumptions.

REMARK 4.1. *It is worth emphasizing that the final steps of scaling and shifting can be avoided if we are only interested, as is customary in network science, in the rankings of the nodes rather than their actual centrality scores.*

5. Block matrix interpretations. Theorems 2.1 and 3.1 show that, loosely, for exponential-based centrality the operation of eliminating backtracking walks is equivalent to replacing the original adjacency matrix A with a suitable block matrix. In the undirected case, the 2 by 2 version Y in (2.2) is used, whereas the directed case has the 3 by 3 version Z in (3.2). These block matrices were constructed indirectly, as a consequence of the generating function approach, [46]. In this section, we add insight by showing how these block matrices may be interpreted directly. We also

show that they remain relevant when the exponential is replaced by a general matrix function.

We begin with an algebraic interpretation. The matrices Z and Y are the *first companion matrix linearization* [18] of the matrix polynomials $(A - S) + (D - I)\lambda - A\lambda^2 + I\lambda^3$ and, respectively, $(D - I) - A\lambda + I\lambda^2$. These are in turn the reversals of, respectively, the directed and undirected version of the deformed graph Laplacian, a matrix polynomial that plays a crucial role in the theory of NBTW combinatorics, see [6, 20]. Algebraically, the companion matrix is the multiplication-by- λI operator in the module obtained by forming quotients in the ring of matrix polynomials using the left ideal generated by the linearized matrix polynomial [34]. Linear-algebraic functions of the companion matrix also have an analogous interpretation; for example, $\exp(tZ)$ represents multiplication by $\exp(t\lambda)I$ modulo $(A - S) + (D - I)\lambda - A\lambda^2 + I\lambda^3$. For more details on this algebraic view of companion matrices see, e.g., [33, Sec. 2], [34, Sec. 5], [35, Sec. 9], and the references therein.

There is also a graph-theoretical interpretation that we now describe in detail. We focus first on the directed case, involving Z in (3.2). A useful connection is that Z represents the adjacency matrix for a three layer weighted network. This network is node-aligned [27], so that the general i th nodes from each layer may be regarded as copies of the same node. Looking at the diagonal blocks of Z , we see that the within-layer connections correspond to the empty graph for layers one and two, and the original graph for layer three. Next, we consider the off-diagonal blocks. The identity matrix in the (1, 2) block and zero matrix in the (1, 3) block show that the only edge out of a node in layer one points to the corresponding node in layer two. Similarly, we may only leave layer two by moving to the corresponding node in layer three. Within layer three, in addition to moving within the layer using edges from A , traversals are possible to layers two and one. The (3, 2) block shows that for $D_{ii} > 1$ a traversal may move from node i in layer three to node i in layer two, with the negative weight $1 - D_{ii}$ assigned to that edge. The (3, 1) block shows that a move from node i in layer three to node j in layer one is allowed if and only if the original graph has a nonreciprocal edge from node i to node j , in which case the inter-layer edge carries weight -1 .

The expression for \mathbf{b} in Theorem 3.1 shows that exponential NBTW centrality arises from a post-processed version of the “standard” exponential centrality on this multilayer network, where negative inter-layer weights have the effect of negating the contributions of backtracking walks in the underlying graph.

The two lemmas and the theorem below show that Z is also relevant for more general matrix functions.

LEMMA 5.1. *For all $r \geq 0$ we have*

$$Z^r \tilde{Z} \begin{bmatrix} 0 \\ 0 \\ I \end{bmatrix} = \begin{bmatrix} p_r(A) \\ p_{r+1}(A) \\ p_{r+2}(A) \end{bmatrix}.$$

Proof. For $r < 2$, the identity may be verified directly. For $r \geq 2$ the result follows by straightforward induction using the three-term recurrence (3.1). \square

THEOREM 5.2 (Directed graph). *Let $\{\alpha_r\}_{r=0}^\infty$ be a sequence such that $f_0(Z) = \sum_{r=0}^\infty \alpha_r Z^r$ converges. Then $\sum_{r=0}^\infty \alpha_r p_r(A)$ converges, and*

$$\sum_{r=0}^\infty \alpha_r p_r(A) = [I \ 0 \ 0] f_0(Z) \begin{bmatrix} I \\ A \\ A^2 - D \end{bmatrix}.$$

Moreover, $\sum_{r=0}^{\infty} \alpha_r p_r(A)$ is equal to the $(3, 3)$ block in $f_0(Z) - f_2(Z)$, where

$$f_k(Z) = \sum_{r=0}^{\infty} \alpha_{r+k} Z^r.$$

Proof. By Lemma 5.1 it follows that

$$\sum_{r=0}^{\infty} \alpha_r p_r(A) = [I \ 0 \ 0] \sum_{r=0}^{\infty} \alpha_r Z^r \tilde{Z} \begin{bmatrix} 0 \\ 0 \\ I \end{bmatrix}$$

and, since

$$\tilde{Z} \begin{bmatrix} 0 \\ 0 \\ I \end{bmatrix} = \begin{bmatrix} I \\ A \\ A^2 - D \end{bmatrix}$$

the first part of the statement follows.

For the second part, note first that $\sum_{r=0}^{\infty} \alpha_{r+2} Z^r$ also converges under the assumptions in the statement. Then

$$\begin{aligned} \sum_{r=0}^{\infty} \alpha_r p_r(A) &= \alpha_0 I + \alpha_1 A + \sum_{r=2}^{\infty} \alpha_r p_r(A) \\ &= \alpha_0 I + \alpha_1 A + [0 \ 0 \ I] \sum_{r=2}^{\infty} \alpha_r Z^{r-2} \tilde{Z} \begin{bmatrix} 0 \\ 0 \\ I \end{bmatrix} \\ &= \alpha_0 I + \alpha_1 A + [0 \ 0 \ I] \sum_{r=2}^{\infty} \alpha_r Z^r \begin{bmatrix} 0 \\ 0 \\ I \end{bmatrix} - [0 \ 0 \ I] \sum_{r=2}^{\infty} \alpha_r Z^{r-2} \begin{bmatrix} 0 \\ 0 \\ I \end{bmatrix} \\ &= [0 \ 0 \ I] \sum_{r=0}^{\infty} (\alpha_r - \alpha_{r+2}) Z^r \begin{bmatrix} 0 \\ 0 \\ I \end{bmatrix}, \end{aligned}$$

as required. \square

Theorem 5.2 shows that we can accumulate combinations of NBTW counts in the original network by looking at the $(3, 3)$ block of an appropriate function of the matrix Z . From a multilayer perspective, the $(3, 3)$ block of Z is associated with the layer containing the adjacency matrix A of the original network. Theorem 5.2 therefore concerns walks of any length that start and end at that layer. This includes not only walks that take place within layer three, but also those that make one or more intermediate visits to the other layers.

The contribution coming from $f_0(Z)$ counts walks in the multilayer network using the original weights. The correction term involving $f_2(Z)$ removes walks from the count by weighting walks of length $r - 2$ as if they were of length r .

In more detail, the term $f_2(Z)$ is used to correct the appearance of an identity matrix in Z^2 that is then propagated at higher powers of Z . This identity matrix at level $r = 2$ originates from the one appearing in the $(3, 2)$ block. This latter is used in the recursion (3.1) defining $p_r(A)$ when $r > 2$ to correct the penalization of walks of the type

$$i \rightarrow \dots \rightarrow k \rightarrow j \rightarrow k \rightarrow j \quad \text{of length } r$$

which are backtracking at length $(r-1)^*$ and thus need not be removed by the factor $p_{r-2}(A)D$. Walks of this type take place entirely in the third layer and are balanced by the existence of walks of the form

$$\begin{array}{cccccccc} \text{node:} & i & \rightarrow & \cdots & \rightarrow & k & \rightarrow & j & \rightarrow & j & \rightarrow & j \\ \text{layer:} & [3] & \rightarrow & \cdots & \rightarrow & [3] & \rightarrow & [3] & \rightarrow & [2] & \rightarrow & [3] \end{array}$$

provided that $r > 2$. When $r = 2$, walks of the form

$$\begin{array}{cccc} \text{node:} & j & \xrightarrow{1-d_j} & j & \xrightarrow{1} & j \\ \text{layer:} & [3] & \rightarrow & [2] & \rightarrow & [3] \end{array}$$

are improperly weighted, as the first hop contains a $+1$ that is used to balance the removal of walks that have never even existed. This introduces a “delay” that is then propagated when longer walks are counted. To counteract it, the matrix function $f_2(Z)$ removes walks of length $r-2$ that are however weighted as if they were of length r ; this type of weighting precisely targets those walks obtained from the propagation of the identity matrix introduced at $r = 2$ in $f_0(Z)$.

Theorem 5.2 can be used to recover the known characterisation of the ordinary generating function. To see this, we let $\alpha_r = t^r$ for all r , with $|t| < \rho(Z)^{-1}$. Then we have

$$\begin{aligned} \sum_{r=0}^{\infty} t^r p_r(A) &= [0 \ 0 \ I] [(I - tZ)^{-1} - t^2(I - tZ)^{-1}] \begin{bmatrix} 0 \\ 0 \\ I \end{bmatrix} \\ &= (1 - t^2)[0 \ 0 \ I](I - tZ)^{-1} \begin{bmatrix} 0 \\ 0 \\ I \end{bmatrix}. \end{aligned}$$

This is equivalent to the result in [6, Section 3].

Similarly, setting $\alpha_r = t^r/r!$ for some $t > 0$, Theorem 5.2 gives

$$(5.1) \quad \sum_{r=0}^{\infty} \frac{t^r p_r(A)}{r!} = [0 \ 0 \ I] (e^{tZ} - t^2 \psi_2(tZ)) \begin{bmatrix} 0 \\ 0 \\ I \end{bmatrix},$$

where $\psi_2(x) = \sum_{r=0}^{\infty} x^r/(r+2)!$. This description of $F(t)$ from (1.1) is equivalent to that given in Theorem 3.1. However, (5.1) casts the result entirely in terms of the (3, 3) block, which, as we have argued above, has a natural multilayer interpretation.

We finish this section by briefly discussing the undirected case. Here, the block matrix Y in (2.2) may be associated with a two-layer network. In layer one, the only possible transition is to the equivalent node in layer two. We may move within layer two using any edge that exists in the original graph, or, for $D_{ii} > 1$, we may move from node i in layer two to node i in layer one along an edge with negative weight $1 - D_{ii}$. The analogues of Lemma 5.1 and Theorem 5.2 are given below, with proofs following in a similar manner.

LEMMA 5.3. *For all $r \geq 0$ we have*

$$Y^r \tilde{Y} \begin{bmatrix} 0 \\ I \end{bmatrix} = \begin{bmatrix} p_{r+1}(A) \\ p_{r+2}(A) \end{bmatrix}.$$

*The additional requirement of the edge $k \rightarrow j$ to be reciprocated is the one that adds the term $p_{r-3}(A)(S - A)$ to the count.

THEOREM 5.4 (Undirected graph). *Let $\{\alpha_r\}_{r=0}^\infty$ be a sequence of nonnegative weights such that $f_k(Y) := \sum_{r=0}^\infty \alpha_{r+k} Y^r$ converges for $k = 0$. Then*

$$\sum_{r=0}^\infty \alpha_r p_r(A) = [0 \ I] (f_0(Y) - f_2(Y)) \begin{bmatrix} 0 \\ I \end{bmatrix}.$$

Theorem 5.4 quantifies in the undirected case how NBTWs on the original network may also be counted by considering walks on the associated multilayer network that start and end in the layer containing A .

6. Star graph analysis. In this section, we give further insights into the effect of restricting to NBTWs in a total communicability centrality measure. For eigenvector centrality, the authors in [29] argued that non-backtracking is a means to avoid *localization*; that is, the concentration of weight on a small number of nodes. Following [19], they characterized localization through an asymptotic limit. Given $\mathbf{v} = (v_i) \in \mathbb{R}^n$ the *inverse participation ratio* is defined as

$$(6.1) \quad \mathcal{S}(\mathbf{v}) = \frac{1}{\|\mathbf{v}\|_2^4} \sum_{i=1}^n v_i^4 = \frac{\|\mathbf{v}\|_4^4}{\|\mathbf{v}\|_2^4}.$$

This quantity is of order one in the case where a small number of elements in \mathbf{v} are responsible for most of the weight. More formally, we say that the measure is *localized* if $\mathcal{S}(\mathbf{v}) = O(1)$ and *nonlocalized* if $\mathcal{S}(\mathbf{v}) = o(1)$, in the asymptotic limit $n \rightarrow \infty$.

We will analyze this property on the undirected star graph with n nodes. Here the adjacency matrix takes the form

$$(6.2) \quad A = \left[\begin{array}{c|c} & \mathbf{1}_{n-1}^T \\ \hline \mathbf{1}_{n-1} & \end{array} \right],$$

where $\mathbf{1}_s \in \mathbb{R}^s$ is the vector of all ones. Thus node 1 is a hub with $n - 1$ connections to nodes $2, 3, \dots, n$, which are only connected to node 1. Intuitively, the effect of backtracking should be significant in this example. The hub node has $n - 1$ neighbours, and hence $n - 1$ walks of length 1, but cannot initiate walks of length two or more without backtracking. Each peripheral node, in contrast, has a single neighbour, but $n - 2$ NBTWs of length two. Hence, in the non-backtracking regime, if t is taken sufficiently large that degree does not dominate, then having one high quality link (to a hub) may be comparable in importance to having many low quality links (to leaves).

It is straightforward to show for (6.2) that (see, for example, [5, 21])

$$e^{tA} = \left[\begin{array}{c|c} \cosh(t\sqrt{n-1}) & \sinh(t\sqrt{n-1})(\sqrt{n-1})^{-1} \mathbf{1}_{n-1}^T \\ \hline \sinh(t\sqrt{n-1})(\sqrt{n-1})^{-1} \mathbf{1}_{n-1} & \cosh(t\sqrt{\mathbf{1}_{n-1} \mathbf{1}_{n-1}^T}) \end{array} \right].$$

Hence, the standard, backtracking walk version of total communicability, $\mathbf{x} = e^{tA} \mathbf{1}$, satisfies

$$\mathbf{x} = \left[\begin{array}{c} \cosh(t\sqrt{n-1}) + \sqrt{n-1} \sinh(t\sqrt{n-1}) \\ \cosh(t\sqrt{n-1}) \mathbf{1}_{n-1} + (\sqrt{n-1})^{-1} \sinh(t\sqrt{n-1}) \mathbf{1}_{n-1} \end{array} \right].$$

So the values x_1 for the hub node and x_2 for a typical leaf node satisfy

$$\begin{aligned} x_1 &= \cosh(t\sqrt{n-1}) + \sqrt{n-1} \sinh(t\sqrt{n-1}), \\ x_2 &= \cosh(t\sqrt{n-1}) + (\sqrt{n-1})^{-1} \sinh(t\sqrt{n-1}). \end{aligned}$$

It follows immediately that for any $t > 0$ we have $x_1 > x_2$ when the graph has at least three nodes. Hence, the standard total communicability measure always ranks the hub node above the leaves. We also find that, for fixed $t > 0$,

$$\begin{aligned} (x_1)^4 &= \frac{n^2}{16} e^{4t\sqrt{n-1}} + o\left(n^2 e^{4t\sqrt{n-1}}\right), \\ (x_i)^4 &= \frac{n^2}{16(n-1)^2} e^{4t\sqrt{n-1}} + o\left(e^{4t\sqrt{n-1}}\right), \quad \text{for all } i = 2, \dots, n, \\ \|\mathbf{x}\|_2^4 &= \frac{n^2}{4} e^{4t\sqrt{n-1}} + o\left(n^2 e^{4t\sqrt{n-1}}\right). \end{aligned}$$

It follows that the inverse participation ratio (6.1) satisfies

$$\lim_{n \rightarrow \infty} \mathcal{S}(\mathbf{x}) = \frac{1}{4}$$

and hence the measure is localized.

We may evaluate the non-backtracking version, $\mathbf{b} = F(t)\mathbf{1}$, from first principles. The hub node has $n-1$ NBTWs of length one and no others, and each leaf node has one NBTW of length one, $n-2$ NBTWs of length two and no others. So

$$b_1 = 1 + t(n-1) \quad \text{and} \quad b_2 = 1 + t + \frac{t^2}{2}(n-2).$$

Straightforward computation shows that $b_1 > b_2$ for $0 < t < 2$ and $b_1 < b_2$ for $t > 2$. Therefore, we observe a transition that is not present with standard exponential centrality, albeit in a parameter regime where t is greater than unity.

More tellingly, for fixed $t > 0$

$$\begin{aligned} (b_1)^4 &= t^4 n^4 + o(n^4), \\ (b_i)^4 &= \frac{t^8}{16} n^4 + o(n^4), \quad \text{for all } i = 2, \dots, n, \\ \|\mathbf{b}\|_2^4 &= \frac{t^8}{16} n^6 + o(n^6). \end{aligned}$$

Hence,

$$\lim_{n \rightarrow \infty} \mathcal{S}(\mathbf{b}) = 0;$$

so the measure is nonlocalized.

7. Numerical tests. In this section we study the localization effect in a realistic setting. We used a random graph model rather than a fixed data set so that we can control the dimension, n , and study the asymptotics. We used an implementation of the Barabási-Albert preferential attachment model for undirected networks in the MATLAB toolbox CONTEST [43]. In this model, nodes are added sequentially until the desired size is reached. Each node is given, upon arrival, d links to current nodes.

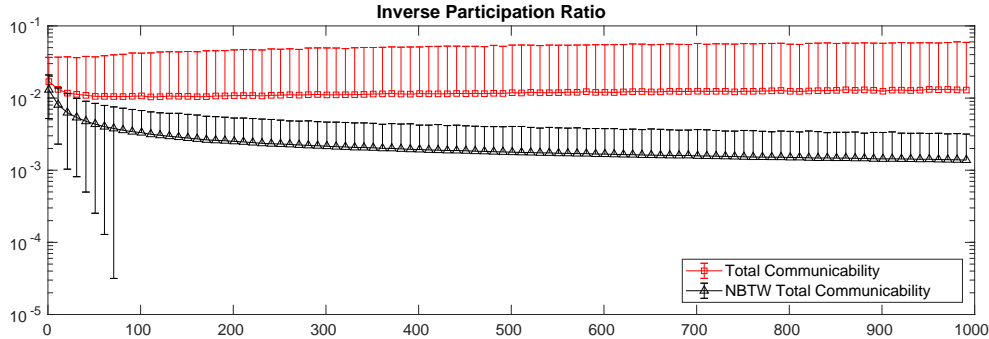


FIG. 1. Inverse participation ratio of vectors $F(t)\mathbf{1}$ and $e^{tA}\mathbf{1}$ as the number of nodes, n , varies.

These target nodes are chosen independently with a probability proportional to their current degree. The resulting network will have a scale-free degree distribution, where a few nodes (*hubs*) have a high degree and many nodes have a very small degree. In this case, the vector of degrees is localized by construction. We have $\rho(A) \leq \|A\|_\infty$, and when this inequality is sharp the presence of a high degree node forces Katz to use a small parameter α , producing results close to degree centrality [8].

We used the default setting of the toolbox, giving each node $d = 2$ links upon arrival to the network, and we built networks of increasing size from $n = 10$ to $n = 1000$. For each network, we computed the inverse participation ratio (6.1) for the normalized total communicability vector $e^{tA}\mathbf{1}$ and its nonbacktracking analogue $\mathbf{b} = F(t)\mathbf{1}$ for $t = 1$ fixed. In order to compute both centrality vectors, we used the MATLAB toolbox `funm_kryl`[†] [1], which computes the action of a matrix function over a vector using a restarted Arnoldi algorithm. We repeated our tests 1000 times.

The results are displayed in Figure 1, where the sample average of the inverse participation ratio is plotted against n in a semilogarithmic scale. The error bar is used to display the standard error around the average values. The results are in line with the analytical results presented in Section 6—the inverse participation ratio decays towards zero for the non-backtracking version, but appears to be bounded away from zero when backtracking walks are included.

8. Summary.

- The main contributions of this work were
1. to derive analytical expressions for the non-backtracking walk exponential generating function for both undirected and directed graphs,
 2. to show that it is feasible to compute the resulting network centrality measures: the key matrix involved in the computations has twice (three times) the dimension of the analogous backtracking version for undirected (directed) graphs, and has a comparable level of sparsity,
 3. to interpret the new measures in the context of traditional centrality on a multilayer network,
 4. to use insights from the multilayer/block matrix setting in order to derive expressions for general matrix function-based centrality measures,
 5. to study how non-backtracking within exponential centrality reduces localization effects on a star graph,
 6. to give further illustrative results on a controllable test network.

[†]Downloaded from <http://guettel.com>

Our overall message is that it is both analytically and computationally attractive to study non-backtracking walk counts in a general matrix function setting that includes the matrix exponential. We therefore hope to have paved the way for future work that, for example, considers

- physical interpretations of the new measures,
- applications to specific fields of network science,
- further links with areas of discrete mathematics, stochastic processes and quantum physics,
- the development and analysis of customized numerical methods for approximating the action of a matrix function on the type of large, sparse unstructured matrix arising in network science.

Acknowledgements. The work of FA and DJH was supported by grant EP/M00158X/1 from the EPSRC/RCUK Digital Economy Programme.

REFERENCES

- [1] Martin Afanasjew, Michael Eiermann, Oliver G Ernst, and Stefan Güttel. Implementation of a restarted Krylov subspace method for the evaluation of matrix functions. *Linear Algebra and its applications*, 429(10):2293–2314, 2008.
- [2] A. H. Al-Mohy and N. J. Higham. Computing the action of the matrix exponential, with an application to exponential integrators. *SIAM J. Sci. Comp.*, 33:488–511, 2011.
- [3] Noga Alon, Itai Benjamini, Eyal Lubetzky, and Sasha Sodin. Non-backtracking random walks mix faster. *Communications in Contemporary Mathematics*, 09:585–603, 2007.
- [4] 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.
- [5] Francesca Arrigo, Michele Benzi, and Caterina Fenu. Computation of generalized matrix functions. *SIAM Journal on Matrix Analysis and Applications*, 37(3):836–860, 2016.
- [6] Francesca Arrigo, Peter Grindrod, Desmond J. Higham, and Vanni Noferini. Non-backtracking walk centrality for directed networks. *Journal of Complex Networks*, (published online ahead of print), 2017.
- [7] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–12, 1999.
- [8] M. Benzi and C. Klymko. On the limiting behavior of parameter-dependent network centrality measures. *SIAM J. Matrix Anal. Appl.*, 36:686–706, 2015.
- [9] Michele Benzi and Christine Klymko. Total communicability as a centrality measure. *Journal of Complex Networks*, 1(2):124–149, 2013.
- [10] P. Bonacich. Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology*, 2:113–120, 1972.
- [11] P. Bonacich. Power and centrality: a family of measures. *American Journal of Sociology*, 92:1170–1182, 1987.
- [12] R. Bowen and O. E. Lanford. Zeta functions of restrictions of the shift transformation. In Shiing-Shen Chern and Stephen Smale, editors, *Global Analysis: Proceedings of the Symposium in Pure Mathematics of the American Mathematical Society, University of California, Berkeley, 1968*, pages 43–49. American Mathematical Society, 1970.
- [13] D. Cvetković, P. Rowlinson, and S. Simić. *Eigenspaces of Graphs*. Cambridge University Press, Cambridge, 1997.
- [14] E. Estrada. *The Structure of Complex Networks*. Oxford University Press, Oxford, 2011.
- [15] Ernesto Estrada and Naomichi Hatano. Communicability in complex networks. *Phys. Rev. E*, 77(3):036111, 2008.
- [16] Ernesto Estrada, Naomichi Hatano, and Michele Benzi. The physics of communicability in complex networks. *Physics Reports*, 514:89–119, 2011.
- [17] Ernesto Estrada and Desmond J. Higham. Network properties revealed through matrix functions. *SIAM Review*, 52:696–671, 2010.
- [18] I. Gohberg, P. Lancaster, and L. Rodman. *Matrix Polynomials*. SIAM, Philadelphia, PA, 2009.
- [19] A. V. Goltsev, S. N. Dorogovtsev, J. G. Oliveira, and J. F. F. Mendes. Localization and spreading of diseases in complex networks. *Phys. Rev. Lett.*, 109:128702, 2012.
- [20] Peter Grindrod, Desmond J. Higham, and Vanni Noferini. The deformed graph Laplacian and

- its applications to network centrality analysis. University of Essex Research Repository 18919, University of Essex, UK, 2017.
- [21] Nicholas J. Higham. *Functions of Matrices: Theory and Computation*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2008.
- [22] Nicholas J. Higham and Edvin Deadman. A catalogue of software for matrix functions. Version 2.0. MIMS EPrint 2016.3, Manchester Institute for Mathematical Sciences, The University of Manchester, UK, January 2016. Updated March 2016.
- [23] Matthew D. Horton. Ihara zeta functions on digraphs. *Linear Algebra and its Applications*, 425:130–142, 2007.
- [24] Matthew D. Horton, H. M. Stark, and Audrey A. Terras. What are zeta functions of graphs and what are they good for? In G. Bertolaiko, R. Carlson, S. A. Fulling, and P. Kuchment, editors, *Quantum graphs and their applications*, volume 415 of *Contemp. Math.*, pages 173–190. 2006.
- [25] L. Katz. A new index derived from sociometric data analysis. *Psychometrika*, 18:39–43, 1953.
- [26] Tatsuro Kawamoto. Localized eigenvectors of the non-backtracking matrix. *Journal of Statistical Mechanics: Theory and Experiment*, 2016:023404, 2016.
- [27] Mikko Kivelä, Alex Arenas, Marc Barthélemy, James P. Gleeson, Yamir Moreno, and Mason A. Porter. Multilayer networks. *Journal of Complex Networks*, 2:203–271, 2014.
- [28] Florent Krzakala, Cristopher Moore, Elchanan Mossel, Joe Neeman, Allan Sly, Lenka Zdeborová, and Pan Zhang. Spectral redemption: clustering sparse networks. *Proceedings of the National Academy of Sciences*, 110:20935–20940, 2013.
- [29] Travis Martin, Xiao Zhang, and M. E. J. Newman. Localization and centrality in networks. *Phys. Rev. E*, 90:052808, 2014.
- [30] Pierre-André G. Maugis, Sofia C. Olhede, and Patrick J. Wolfe. Topology reveals universal features for network comparison. *arXiv:1705.05677 [stat.ME]*, 2017.
- [31] Flaviano Morone and Hernán A. Makse. Influence maximization in complex networks through optimal percolation. *Nature*, 524:65–68, 2015.
- [32] Flaviano Morone, Byungjoon Min, Lin Bo, Romain Mari, and Hernán A. Makse. Collective influence algorithm to find influencers via optimal percolation in massively large social media. *Scientific Reports*, 6:30062, 2016.
- [33] Y. Nakatsukasa and V. Noferini. On the stability of computing polynomial roots via confederate linearizations. *Mathematics of Computation*, 85:2391–2425, 2016.
- [34] 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.
- [35] V. Noferini and F. Poloni. Duality of matrix pencils, Wong chains and linearizations. *Linear Algebra Appl.*, 471:730–767, 2015.
- [36] Romualdo Pastor-Satorras and Claudio Castellano. Distinct types of eigenvector localization in networks. *Scientific Reports*, 6:18847, 2016.
- [37] Yousef Saad. Analysis of some Krylov subspace approximations to the matrix exponential operator. *SIAM Journal on Numerical Analysis*, 29(1):209–228, 1992.
- [38] Alaa Saade, Florent Krzakala, and Lenka 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.
- [39] Uzy Smilansky. Quantum chaos on discrete graphs. *Journal of Physics A: Mathematical and Theoretical*, 40:F621, 2007.
- [40] Sasha Sodin. Random matrices, non-backtracking walks, and the orthogonal polynomials. *J. Math. Phys.*, 48:123503, 2007.
- [41] H.M. Stark and A.A. Terras. Zeta functions of finite graphs and coverings. *Advances in Mathematics*, 121(1):124–165, 1996.
- [42] Andrei Tarfulea and Robert Perlis. An Ihara formula for partially directed graphs. *Linear Algebra and its Applications*, 431:73–85, 2009.
- [43] Alan Taylor and Desmond J. Higham. CONTEST: A controllable test matrix toolbox for MATLAB. *ACM Trans. Math. Softw.*, 35(4):26:1–26:17, 2009.
- [44] Audrey Terras. *Harmonic Analysis on Symmetric Spaces — Euclidean Space, the Sphere, and the Poincaré Upper Half-Plane*. Springer, New York, 2nd edition, 2013.
- [45] Yusuke Watanabe and Kenji 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.
- [46] Herbert S. Wilf. *Generatingfunctionology*. A. K. Peters, Ltd., Natick, MA, USA, 2006.