



# Non-backtracking PageRank

Francesca Arrigo<sup>1</sup> · Desmond J. Higham<sup>2</sup> · Vanni Noferini<sup>3</sup>

Received: 14 December 2018 / Revised: 17 May 2019 / Accepted: 21 May 2019  
© The Author(s) 2019

## Abstract

The PageRank algorithm, which has been “bringing order to the web” for more than 20 years, computes the steady state of a classical random walk plus teleporting. Here we consider a variation of PageRank that uses a non-backtracking random walk. To do this, we first reformulate PageRank in terms of the associated line graph. A non-backtracking analog then emerges naturally. Comparing the resulting steady states, we find that, even for undirected graphs, non-backtracking generally leads to a different ranking of the nodes. We then focus on computational issues, deriving an explicit representation of the new algorithm that can exploit structure and sparsity in the underlying network. Finally, we assess effectiveness and efficiency of this approach on some real-world networks.

**Keywords** PageRank · Non-backtracking · Complex networks · Centrality

**Mathematics Subject Classification** 05C50 · 05C82 · 65F10 · 65F50

## 1 Motivation

The notion of a walk around a graph is both natural and useful. The walker may follow any available edge, with nodes and edges being revisited at any stage. However, in some settings, walks that *backtrack*—leave a node and then return to it on the next step—are best avoided. The concept of restricting attention to *non-backtracking walks* has arisen,

---

The work of FA was supported by fellowship ECF-2018-453 from the Leverhulme Trust. The work of DJH was supported by EPSRC/RCUK Established Career Fellowship EP/M00158X/1 and by EPSRC Programme Grant EP/P020720/1. EPSRC data statement: the data used in this work is publicly available from [32].

---

✉ Francesca Arrigo  
francesca.arrigo@strath.ac.uk

Desmond J. Higham  
d.j.higham@ed.ac.uk

Vanni Noferini  
vanni.noferini@aalto.fi

<sup>1</sup> University of Strathclyde, Glasgow, UK

<sup>2</sup> University of Edinburgh, Edinburgh, UK

<sup>3</sup> Aalto University, Aalto, Finland

essentially independently, in a wide range of seemingly unrelated fields, including spectral graph theory [2,13,14], number theory [30], discrete mathematics [4,28], quantum chaos [26], random matrix theory [27], stochastic analysis [1], applied linear algebra [29] and computer science [24,33]. Non-backtracking has more recently been considered in the context of matrix computation, where it has been shown to form the basis of effective algorithms in network science for community detection, centrality and alignment [3,10,16,18,20,21,23,31]. We note that non-backtracking walks have typically been studied on undirected networks, but the definition continues to make sense in the directed case.

For network science applications, non-backtracking can be imposed at very little extra cost and offers quantifiable benefits: centrality measures can avoid localization [3,10,20] and discover key influencers [21,31], and community detection algorithms perform optimally on the stochastic block model [18]. In terms of classical random walks on a graph, it is also known that non-backtracking enhances mixing [1].

Given that the PageRank algorithm [22] computes the steady state of a particular (teleporting) random walk, it is therefore natural to ask whether there is any scope for designing and analysing a non-backtracking analog. We emphasize that the PageRank philosophy has been extended well beyond the WWW into a diverse range of areas including bioinformatics, bibliometrics, business, chemistry, neuroscience, social media, and sports [9], where non-backtracking may be a natural requirement. In particular, PageRank-like measures have been extensively used in transport modelling; see, for example, [6,25]. In this setting it is reasonable to assume that a vehicle will not immediately head back to the same road junction, train station or bus stop.

In this work, we therefore consider the following general questions:

- What is an appropriate non-backtracking version of PageRank?
- Can the imposition of non-backtracking lead to a different node ranking?
- Is there an efficient way to implement such an algorithm?

At the core of our analysis is the Hashimoto matrix, which arises when we work on the *dual graph*, or *line graph*. It is interesting to note a connection with the *space syntax* approach in the analysis of road networks, where edges in a graph represent crossroads and nodes are the streets [7]. This representation is the dual of the more intuitive setting where edges represent roads and nodes are intersections; however, in [25] the authors argue that “the dual graphs carry more information than the corresponding primal graph.”

The material is organized as follows. In Sect. 2 we describe background material on PageRank. Section 3 gives a transposition of standard PageRank to the dual space and verifies the equivalence of the two formulations. Section 4 defines and analyzes a new non-backtracking analog. In particular, we show that, even for undirected graphs, non-backtracking generally leads to a different ranking of the nodes. In Sect. 5 we derive a computable description of the new measure and show how to exploit the structure and sparsity of the Hashimoto matrix. In Sect. 6 we comment on some numerical experiments performed on road networks. Section 7 gives some conclusions.

## 2 Background

Let  $A = (a_{ij}) \in \mathbb{R}^{n \times n}$  be the adjacency matrix of an unweighted, weakly connected, directed graph with  $n$  nodes and  $\tilde{m}$  directed edges. We further assume that there are no multiple edges linking nodes in the network. Under these assumptions,  $a_{ij} = 1$  if there is an edge from node  $i$  to node  $j$ , otherwise  $a_{ij} = 0$ . A node that has no outgoing links is called a *dangling*

node. We use the common approach of removing dangling nodes by modifying the adjacency matrix; all-zeros rows in  $A$  are replaced by all-ones rows [15]. This corresponds to adding  $n$  edges from each dangling node to all the nodes in the network, itself included.

Let  $G = (V, E)$  be the graph obtained once these edges have been added to the original network in order to remove dangling nodes. In the remainder of the paper, we shall work on this network. Let  $W = A + \chi \mathbf{1}^T$  be its adjacency matrix, where  $\chi$  is the indicator vector for the set containing the dangling nodes in the original digraph. Here,  $\mathbf{1} \in \mathbb{R}^n$  denotes the vector of all ones, and we will sometimes write  $\mathbf{1}_n$  when its dimension may be in doubt. We let  $k$  be the number of dangling nodes in the original network, so the new graph  $G$  contains  $m := \tilde{m} + kn$  edges among  $n$  nodes.

Let  $\mathbf{d} = (d_i) = W\mathbf{1}$  be the vector of *out-degrees* for the nodes in the graph represented by  $W$ , i.e., let  $d_i$  be the number of edges leaving node  $i$  in  $G$ . We will denote by  $D$  the diagonal matrix whose diagonal entries are  $D_{ii} = d_i$  for all  $i$ . Since all the nodes have positive out-degree,  $D$  is invertible.

We assume that the edges have been labelled in some arbitrary order  $\{e\}_{e=1}^m$ . We also use the notation  $i \rightarrow j$  to denote an edge from node  $i$  to node  $j$  and  $i \rightarrow \star$  to denote an edge that starts from node  $i$ . We then define the matrices  $L, R \in \mathbb{R}^{m \times n}$  as follows:

$$L_{ei} = \begin{cases} 1 & \text{if } e \text{ has the form } i \rightarrow \star \\ 0 & \text{otherwise} \end{cases}$$

and

$$R_{ej} = \begin{cases} 1 & \text{if } e \text{ has the form } \star \rightarrow j \\ 0 & \text{otherwise.} \end{cases}$$

In words  $L_{ei}$  records whether edge  $e$  starts at node  $i$ , and  $R_{ej}$  records whether edge  $e$  ends at node  $j$ . The matrices  $L$  and  $R$  will be referred to as the *source matrix* and *target matrix*, respectively. It is readily seen that  $W = L^T R$  and  $D = L^T L$ . Moreover, the matrix  $R^T R$  is diagonal with the in-degree of the nodes as its diagonal entries.

**Remark 1** If  $W = A$  and the original network contained no dangling nodes, then  $D$  would still be invertible, but  $R^T R$  might not be invertible as the network may contain *source nodes*; that is, nodes with no incoming links. On the other hand, if the original graph contained at least one dangling node, so that  $W \neq A$ , then its removal would also remove any source node, thus making  $R^T R$  invertible as well.

As discussed in the original work [22], the PageRank algorithm may be motivated from two different perspectives: (a) the long-term behavior of a web surfer who randomly follows outgoing links and occasionally teleports (jumps) to a random page, and (b) the outcome of an iterative voting procedure where a page votes for pages that it follows, with more important pages given more influence.

From the random surfer viewpoint, the PageRank vector  $\mathbf{x}$  may be defined as the stationary distribution of a Markov chain that is constructed from a random walk on the graph plus teleporting:

$$P^T \mathbf{x} = \mathbf{x}, \tag{1}$$

where  $P = \alpha D^{-1} W + (1 - \alpha) \mathbf{1} \mathbf{v}^T$  is row stochastic and  $\|\mathbf{x}\|_1 = 1$ , [9]. Here,  $W^T D^{-1}$  is column-stochastic,  $\alpha \in (0, 1)$ , and  $\mathbf{v} \geq \mathbf{0}$  is such that  $\|\mathbf{v}\|_1 = 1$ . The matrix  $F = \mathbf{1} \mathbf{v}^T$  is sometimes referred to as the *teleportation matrix* and  $\mathbf{v}$  the *teleportation distribution vector* [9] or *personalization vector* [19]. The scalar  $\alpha$  is known as the *teleportation parameter* [5,9] or *amplification factor* [34].

Based on the voting procedure, we may equivalently define the PageRank vector as the normalized version,  $\mathbf{x} = \tilde{\mathbf{x}}/\|\tilde{\mathbf{x}}\|_1$ , of the non-negative solution to the following linear system:

$$(I - \alpha W^T D^{-1})\tilde{\mathbf{x}} = (1 - \alpha)\mathbf{v}. \quad (2)$$

To see that these two formulations of the PageRank problem are equivalent, starting from (1) we have

$$\mathbf{x} = P^T \mathbf{x} = \alpha W^T D^{-1} \mathbf{x} + (1 - \alpha)\mathbf{v}(\mathbf{1}^T \mathbf{x}) \quad (3)$$

and (2) follows by rearranging the terms and noting  $(\mathbf{1}^T \mathbf{x}) = 1$ . In our work, the random walk interpretation will be used as the basis for a non-backtracking extension.

In the following, for simplicity, we use the original uniform teleporting distribution vector  $\mathbf{v} = \frac{1}{n}\mathbf{1}$ , so that (1) rewrites as

$$\left( \alpha W^T D^{-1} + \frac{(1 - \alpha)}{n} \mathbf{1}\mathbf{1}^T \right) \mathbf{x} = \mathbf{x}, \quad (4)$$

and (2) rewrites as

$$(I - \alpha W^T D^{-1})\tilde{\mathbf{x}} = \frac{(1 - \alpha)}{n} \mathbf{1}. \quad (5)$$

Let us briefly recall here that if we let  $D_A$  be the diagonal matrix whose diagonal entries are  $(D_A)_{ii} = (A\mathbf{1})_i$  and if we use the symbol  $\dagger$  to denote the Moore–Penrose pseudo-inverse of a matrix [12], then the solution to the linear system

$$(I - \alpha A^T D_A^\dagger)\mathbf{x} = \frac{(1 - \alpha)}{n} \mathbf{1} \quad (6)$$

is strictly related to the solution of the *PageRank problem*, i.e., to the solution of (5). Indeed, let us first recall that, since the matrix  $A^T D_A^\dagger \geq 0$  is column-substochastic, i.e., it is such that  $\mathbf{1}^T A^T D_A^\dagger \leq \mathbf{1}^T$ , the problem of finding the non-negative solution to the linear system (6) is usually referred to as the *pseudo-PageRank problem*. Moreover, the following result holds.

**Theorem 1** [8, Theorem 4.2] *The PageRank vector  $\mathbf{x}$  solution to (5) is obtained by normalizing the solution  $\mathbf{x}$  to the pseudo-PageRank problem (6):*

$$\mathbf{x} = \frac{\mathbf{x}}{\|\mathbf{x}\|_1}.$$

This theorem shows that, in order to compute the PageRank vector  $\mathbf{x}$  one can solve the pseudo-PageRank problem and then normalize the solution. Therefore, underlying every pseudo-PageRank system there is a true PageRank problem.

In the following section we show that PageRank can be reformulated in the dual space, where edges in  $G$  play the role of nodes in the new network. This reformulation then leads to a natural non-backtracking extension.

### 3 Moving to the Dual Space

Let us consider the line graph associated with the network represented by  $W$ . We will use a calligraphic font to denote matrices associated with the line graph. In the line graph, nodes are edges in the original network  $G$  and there is a connection between two edges if the endpoint

of the first is the source point of the second. Let  $\mathcal{W} \in \mathbb{R}^{m \times m}$  be the adjacency matrix of the line graph, i.e., the matrix whose entries are:

$$(\mathcal{W})_{i \rightarrow j, k \rightarrow l} = \delta_{jk},$$

where  $\delta_{jk}$  is the Kronecker delta. It is easily proved that  $\mathcal{W} = RL^T$ . Let us define the diagonal matrix  $\mathcal{D}$  whose diagonal entries are the entries of the vector  $\mathcal{W}\mathbf{1}$ . This stores the “out-degree” of each edge in the network, i.e., the number of edges leading out of the current edge. Therefore, an edge  $i \rightarrow j$ , has out-degree in the line graph coinciding with the number of edges leaving node  $j$ , i.e., the out-degree  $d_j$  of node  $j$ . We are assuming that the network does not contain dangling nodes, so  $\mathcal{W}\mathbf{1} > \mathbf{0}$  and  $\mathcal{D}$  is invertible.

Our immediate goal is to define a PageRank vector in the dual space, i.e., the edge space, as the solution of a linear system involving  $\mathcal{W}$  that, once projected, will allow us to retrieve the PageRank vector  $\mathbf{x}$  defined earlier in the primal space. This forms the basis of the non-backtracking version that we develop in Sect. 4.

### 3.1 Formulation of Edge PageRank

We will define the transition matrix for edge PageRank in terms of the matrices  $\mathcal{W}$ ,  $\mathcal{D}$  and an appropriately defined teleportation matrix  $\mathcal{F}$ . The new transition matrix will then have the form:

$$\mathcal{P} = \alpha \mathcal{D}^{-1} \mathcal{W} + (1 - \alpha) \mathcal{F},$$

where we must specify the form of  $\mathcal{F}$ . Some care is needed in the construction of  $\mathcal{F}$  to ensure that the original PageRank vector can be recovered—the essential point is that jumping to an edge at random does not correspond to jumping to a node at random, when the graph is not regular. To proceed, we note that each edge in the original graph may be mapped to the node that it points to. Let  $N$  be the number of nodes in the network that have positive in-degree, i.e., the number of nodes in the network that are not source nodes. We then define our teleportation matrix as

$$\mathcal{F} = \frac{1}{N} \mathbf{1} \boldsymbol{\pi}^T$$

where  $\boldsymbol{\pi} = R(R^T R)^\dagger \mathbf{1}$  is the vector that associates to each edge the inverse of the in-degree of its endpoint. Here the symbol  $\dagger$  denotes the pseudo-inverse of the diagonal matrix  $R^T R$ ; indeed, as pointed out in Remark 1 if the network contains a source, i.e., a node with zero in-degree, then the matrix  $R^T R$  is not invertible. It is worth stressing that, however,  $\boldsymbol{\pi} > \mathbf{0}$  even when the network contains a source, thanks to the premultiplication by  $R$  that is used to assign to each edge the inverse of the in-degree of its endpoint (that is thus positive, since there is at least the particular edge that points to it). Since  $\|\boldsymbol{\pi}\|_1 = N$ , it holds that  $\mathcal{F}\mathbf{1} = \mathbf{1}$ . Moreover, we know that  $\mathcal{D}^{-1} \mathcal{W}\mathbf{1} = \mathbf{1}$ . Therefore, the invariant measure  $\hat{\mathbf{x}} \geq \mathbf{0}$  such that  $\|\hat{\mathbf{x}}\|_1 = 1$  satisfies

$$\mathcal{P}^T \hat{\mathbf{x}} = \hat{\mathbf{x}}.$$

After some algebraic manipulation, the above equation rewrites as

$$(I - \alpha \mathcal{W}^T \mathcal{D}^{-1}) \hat{\mathbf{x}} = (1 - \alpha) \frac{\boldsymbol{\pi}}{\|\boldsymbol{\pi}\|_1}. \tag{7}$$

These two expressions translate PageRank from the node space to its dual. We have not yet shown that these expressions are equivalent to the ones described in Sect. 2. In order to do

so, we first need to describe how to project  $\hat{\mathbf{x}}$  back to the node space. This can be done by pre-multiplying the vector  $\hat{\mathbf{x}}$  by  $R^T$ . The projected non-negative vector  $\mathbf{z} := R^T \hat{\mathbf{x}}$  can be easily seen to have unit 1-norm.

Our goal is now to show that  $\mathbf{z} = \mathbf{x}$ , thus deriving equivalence between the new formulations of PageRank and (4)–(5). Let us first prove two preliminary results.

**Lemma 1** *In the above notation,  $\mathcal{D}^{-1}RD = R$ .*

**Proof** Entrywise it holds

$$(\mathcal{D}^{-1}RD)_{ej} = \begin{cases} \frac{1}{d_j}d_j & \text{if } e = \star \rightarrow j \\ 0 & \text{otherwise} \end{cases} = R_{ej}.$$

□

**Lemma 2** *In the above notation,  $L^T\mathcal{D}^{-1}R = WD^{-1}$ .*

**Proof** Left-multiplying by  $L^T$  the relation in Lemma 1 and using  $W = L^T R$  it follows that  $L^T\mathcal{D}^{-1}RD = W$ , thus the conclusion, since  $D$  is invertible. □

**Theorem 2** *In the above notation, if a network does not contain a source node, then for all  $\alpha \in (0, 1)$  we have*

$$(I - \alpha\mathcal{W}^T\mathcal{D}^{-1})^{-1} = R^T(I - \alpha\mathcal{W}^T\mathcal{D}^{-1})^{-1}R(R^T R)^{-1}.$$

**Proof** The result can be proved by using the facts that  $\mathcal{W}^T = LR^T$  and  $W^T = R^T L$ , and appealing to Lemmas 1 and 2:

$$\begin{aligned} (I - \alpha\mathcal{W}^T\mathcal{D}^{-1})^{-1} &= I + \alpha LR^T\mathcal{D}^{-1} + \alpha^2 L(R^T\mathcal{D}^{-1}L)R^T\mathcal{D}^{-1} + \alpha^3 L(R^T\mathcal{D}^{-1}L)^2 R^T\mathcal{D}^{-1} + \dots \\ &= I + \alpha LR^T\mathcal{D}^{-1} + \alpha^2 L(D^{-1}W^T)R^T\mathcal{D}^{-1} + \alpha^3 L(D^{-1}W^T)^2 R^T\mathcal{D}^{-1} + \dots \\ &= I + \alpha L \left[ I + \alpha D^{-1}W^T + \alpha^2 (D^{-1}W^T)^2 + \dots \right] R^T\mathcal{D}^{-1}, \end{aligned}$$

where the first equality holds because  $\rho(\mathcal{W}^T\mathcal{D}^{-1}) = 1 > \alpha$ , and thus

$$(I - \alpha\mathcal{W}^T\mathcal{D}^{-1})^{-1} = \sum_{k=0}^{\infty} \alpha^k (\mathcal{W}^T\mathcal{D}^{-1})^k.$$

Therefore,

$$\begin{aligned} R^T(I - \alpha\mathcal{W}^T\mathcal{D}^{-1})^{-1}R(R^T R)^{-1} &= I + \alpha R^T L \left[ I + \alpha D^{-1}W^T + \alpha^2 (D^{-1}W^T)^2 + \dots \right] R^T\mathcal{D}^{-1}R(R^T R)^{-1} \\ &= I + \left[ \alpha W^T\mathcal{D}^{-1} + \alpha^2 (W^T\mathcal{D}^{-1})^2 + \dots \right] R^T R(R^T R)^{-1} \end{aligned}$$

and hence the conclusion. □

The following corollary is then immediate.

**Corollary 1** *When the network does not contain a source node, the edge PageRank vector in (7) projects to the original PageRank vector in (3); that is,  $R^T \hat{\mathbf{x}} = \mathbf{x}$ .*

**Proof** Using Theorem 2 and  $\boldsymbol{\pi} = R(R^T R)^{-1}\mathbf{1}$  we have

$$\mathbf{x} = \frac{(1-\alpha)}{n}(I - \alpha W^T D^{-1})^{-1}\mathbf{1} = \frac{(1-\alpha)}{\|\boldsymbol{\pi}\|_1} R^T (I - \alpha \mathcal{W}^T \mathcal{D}^{-1})^{-1}\boldsymbol{\pi} = R^T \hat{\mathbf{x}},$$

where we have used the fact that  $N = n$  and thus  $\|\boldsymbol{\pi}\|_1 = n$ . □

We now describe in detail what happens when the network contains  $s = n - N \geq 1$  source nodes. We will show that the edge and node based systems are not equivalent. It is worth pointing out that we are tacitly assuming that the original network did not contain dangling nodes. If this was not the case, then the removal of the dangling nodes would have also removed the source nodes; see Remark 1.

The adjacency matrix of the network containing  $s$  source nodes, which we assume to be nodes  $1, \dots, s$  without loss of generality, can be described as

$$W = A = \left[ \begin{array}{c|c} 0_{s,s} & S \\ \hline 0_{N,s} & \tilde{W} \end{array} \right],$$

where  $S \in \mathbb{R}^{s \times N}$  describes the edges leaving source nodes to target non-source nodes in the network and  $\tilde{W}$  is the adjacency matrix of the graph obtained by removing nodes  $1, \dots, s$  and all the edges leaving them. Moreover, it holds that

$$D = \left[ \begin{array}{c|c} D_S & \\ \hline & \tilde{D} \end{array} \right],$$

where the diagonal matrix  $D_S$  is such that  $(D_S)_{ii} = (S\mathbf{1})_i$  is the out-degree of the  $i$ th source node, for  $i = 1, \dots, s$  and  $\tilde{D}$  is the diagonal matrix such that  $(\tilde{D})_{ii} = (\tilde{W}\mathbf{1})_i$  for  $i = 1, 2, \dots, N$ . The diagonal matrix  $R^T R$  of in-degrees is no longer invertible, since nodes  $1, \dots, s$  have no incoming links and thus zero in-degree. An easy computation shows that

$$(I - \alpha W^T D^{-1})^{-1} = \left[ \begin{array}{c|c} I_s & 0_{s,N} \\ \hline Y & (I - \alpha \tilde{W}^T \tilde{D}^{-1})^{-1} \end{array} \right]$$

with  $Y = \alpha(I - \alpha \tilde{W}^T \tilde{D}^{-1})^{-1} S^T D_S^{-1}$ . It thus follows that

$$\mathbf{x} = \beta (I - \alpha W^T D^{-1})^{-1}\mathbf{1} = \beta \begin{bmatrix} \mathbf{1}_s \\ \star \end{bmatrix},$$

where  $\beta = \frac{(1-\alpha)}{n}$ .

On the other hand, it is easily checked following a reasoning similar to the one outlined in the proof of Theorem 2 that if we use the matrix defined in the edge space and then project, we obtain:

$$\mathbf{z} = \beta R^T (I - \alpha \mathcal{W}^T \mathcal{D}^{-1})^{-1} R(R^T R)^\dagger \mathbf{1} = \beta \begin{bmatrix} \mathbf{0}_s \\ \star \end{bmatrix},$$

and therefore  $\mathbf{z} \neq \mathbf{x}$  and we cannot extend the statement in Theorem 2 by simply replacing  $(R^T R)^{-1}$  with  $(R^T R)^\dagger$  if the network contains at least one source node.

In order to overcome this minor issue, we use a slightly less intuitive approach and, instead of mapping each edge to its endpoint, we map it to its start point. With this convention, the

matrix  $D = L^T L$  of out-degrees is always invertible, even in the pathological case of a network containing one or more source nodes but no dangling nodes. Using as personalization vector

$$\mathbf{v} = L(L^T L)^{-1} \mathbf{1} = LD^{-1} \mathbf{1}$$

which now has  $\|\mathbf{v}\|_1 = n$ , and projecting with  $L^T$  instead of  $R^T$ , Theorem 2 rewrites as follows.

**Theorem 3** *In the above notation, for all  $\alpha \in (0, 1)$  we have*

$$(I - \alpha W^T D^{-1})^{-1} = L^T (I - \alpha W^T D^{-1})^{-1} L D^{-1}.$$

**Proof** The result can be proved using the fact that  $W^T = LR^T$ , the fact that  $\rho(W^T D^{-1}) = 1 > \alpha$  and Lemma 2:

$$\begin{aligned} & L^T (I - \alpha W^T D^{-1})^{-1} L D^{-1} \\ &= L^T \left[ I + \alpha W^T D^{-1} + \alpha^2 W^T D^{-1} W^T D^{-1} + \dots \right] L D^{-1} \\ &= I + L^T \left[ \alpha L (R^T D^{-1} L) + \alpha^2 L (R^T D^{-1} L)^2 + \dots \right] D^{-1} \\ &= I + L^T \left[ \alpha L D^{-1} W^T + \alpha^2 L D^{-1} W^T D^{-1} W^T + \dots \right] D^{-1} \\ &= I + \alpha W^T D^{-1} + \alpha^2 W^T D^{-1} W^T D^{-1} + \dots \\ &= (I - \alpha W^T D^{-1})^{-1}. \end{aligned}$$

□

We then have the following general corollary.

**Corollary 2** *After the network has been preprocessed to eliminate dangling nodes (if any), the projected solution via  $L$  to PageRank in the edge space*

$$(I - \alpha W^T D^{-1}) \hat{\mathbf{x}} = \frac{(1 - \alpha)}{n} L D^{-1} \mathbf{1}.$$

*is equivalent to standard PageRank described in (5), i.e.,  $L^T \hat{\mathbf{x}} = \mathbf{x}$ .*

**Proof** A proof follows immediately from Theorem 3. □

## 4 The Non-backtracking Framework

In this Section we want to exploit the definition of PageRank in the dual space to describe a PageRank-like system in the non-backtracking framework. Recall that a walk is said to be *backtracking* if it contains at least one instance of  $i \rightarrow j \rightarrow i$ , where  $i, j$  are two nodes in the network connected by reciprocated edges. Let  $\mathcal{B} \in \mathbb{R}^{m \times m}$  be the matrix whose entries are

$$\mathcal{B}_{i \rightarrow j, k \rightarrow \ell} = \delta_{jk} (1 - \delta_{i\ell}).$$

This matrix has a 1 in position  $(i \rightarrow j, k \rightarrow \ell)$  if the two edges are consecutive (i.e., if  $j = k$ ) but they do not form a backtracking walk (i.e., if  $i \neq \ell$ ). We will refer to this matrix as the *non-backtracking edge-matrix* or the *Hashimoto matrix* [11]. If we let  $\circ$  represent



the Schur product, i.e., the entry-wise product, between two matrices, then the Hashimoto matrix can be described as  $\mathcal{B} = \mathcal{W} - \mathcal{W} \circ \mathcal{W}^T$ . Even though this matrix is built after all dangling nodes have been removed from the network, it is still possible for it to have zero rows. This happens when some of the nodes have no outgoing links, apart from exactly one reciprocated edge. Let us assume that there is one node of this type in the network, node  $i$ , connected to node  $j$  through a reciprocated edge. If the random walker reaches node  $i$  using the edge  $j \rightarrow i$ , then it will not be able to leave this node without backtracking; therefore, the row corresponding to the edge  $j \rightarrow i$  only contains zeros. We will refer to edges of this type as *dangling edges*. Thanks to Theorem 1 we do not need to correct for the presence of these edges in the Hashimoto matrix in order to compute the non-backtracking version of PageRank in the edge space. Indeed, a rescaled solution to the pseudo-PageRank problem formulated with  $\mathcal{B}$  will be the solution to the PageRank problem, where the zero rows in the matrix  $\mathcal{B}$  have been replaced with the vector  $\mathbf{v}^T = (LD^{-1}\mathbf{1})^T$ . We note that this correction has the effect of allowing backtracking in this very special circumstance.

In the following we will assume that there are no dangling edges in the graph; the remainder of this section can be adapted to the case where such edges existed.

To finalize the description of the transition matrix that will allow us to compute the non-backtracking version of PageRank, we need to describe the teleportation matrix. Here we must make a choice: should we allow backtracking during a teleportation step? In this work we allow such behavior on the basis that the original PageRank philosophy is to make teleportation independent of the network topology. (Further, we note that teleportation in the original PageRank algorithm allows a step of the form  $i \rightarrow i$  even when the graph has no loops, violating the classical definition of a walk. Hence, teleportation can be regarded as a separate process, independent of the underlying random walk.) We also note that backtracking due to teleportation will be an extremely low probability event in large networks, which are the ones of interest in this context. We therefore use

$$\mathcal{F} = \frac{1}{n}\mathbf{1}\mathbf{v}^T$$

with  $\mathbf{v} = LD^{-1}\mathbf{1}$ . Recall that  $\|\mathbf{v}\|_1 = n$ .

We can now define for  $0 < \alpha < 1$  the matrix  $\mathcal{P} = \alpha\mathcal{D}^{-1}\mathcal{B} + \frac{(1-\alpha)}{n}\mathbf{1}\mathbf{v}^T$ , where now  $\mathcal{D}$  is the diagonal matrix whose diagonal entries are the entries of the vector  $\mathcal{B}\mathbf{1}$ . Since we assume that there are no dangling edges,  $\mathcal{B}\mathbf{1} > 0$  and we have that  $\mathcal{P}\mathbf{1} = \mathbf{1}$ ; therefore, the PageRank vector in the dual space will be the unit 1-norm, non-negative vector  $\hat{\mathbf{y}}$  that solves the following linear system

$$(I - \alpha\mathcal{B}^T\mathcal{D}^{-1})\hat{\mathbf{y}} = \frac{(1-\alpha)}{n}\mathbf{v}. \tag{8}$$

In order to then derive a nodal measure, we need to project  $\hat{\mathbf{y}}$  onto the node space; following Theorem 3, we use the matrix  $L^T$ :

$$\mathbf{y} := L^T\hat{\mathbf{y}}. \tag{9}$$

This vector will be non-negative and will have unit 1-norm, since  $L\mathbf{1} = \mathbf{1}$  and  $\|\hat{\mathbf{y}}\|_1 = 1$ . Therefore, we can define the non-backtracking PageRank-like vector as follows.

**Definition 1** Let  $\mathcal{B}$  be the Hashimoto matrix of a directed graph with no dangling nodes or dangling edges. Let  $\mathcal{D}$  be the diagonal matrix whose diagonal entries are the entries of the vector  $\mathcal{B}\mathbf{1}$ , and let  $L$  be the source matrix of the graph. For all  $0 < \alpha < 1$  we define the *non-backtracking PageRank-like* vector as the non-negative, unit 1-norm vector  $\mathbf{y} := L^T\hat{\mathbf{y}}$ , where  $\hat{\mathbf{y}} \geq 0$  with  $\|\hat{\mathbf{y}}\|_1 = 1$  solves the the linear system (8), for  $\mathbf{v} = L(L^T L)^{-1}\mathbf{1}$ .

Before commenting on the features of this new measure, let us point out that the new PageRank problem (8) is equivalent to the eigenproblem

$$\mathcal{P}^T \hat{\mathbf{y}} = \hat{\mathbf{y}},$$

where  $\mathcal{P} = \alpha \mathcal{D}^{-1} \mathcal{B} + \frac{(1-\alpha)}{n} \mathbf{1} \mathbf{v}^T$  and  $\|\hat{\mathbf{y}}\|_1 = 1$ , in line with the equivalent formulations (5) and (4) for standard PageRank.

Definition 1 allows us to introduce the non-backtracking aspect in PageRank. As we will see below, there are situations in which the newly introduced measure and standard PageRank return vectors that are the same or induce the same ranking. However, in applications where backtracking is best avoided, the two measures may differ, as we shall see in Sect. 6.

Let us now briefly discuss the case where the two measures behave similarly. It has been proved in [17] that for an undirected graph with nodes all having at least degree two (so that there are no dangling edges), the steady state of  $P = D^{-1}W$  coincides with that of  $\mathcal{P} = \mathcal{D}^{-1}\mathcal{B}$ , once the latter has been projected in the node space via  $L^T$ : so standard PageRank and the new non-backtracking PageRank-like measures are the same in the absence of teleportation. In other words, in this case where every edge is reciprocated and no teleportation is allowed, this result tells us that the long-term frequency of visits to a particular node during a random walk does not change when non-backtracking is imposed—the nodes benefit from non-backtracking precisely in proportion to their original PageRank score. Intuitively, nodes given higher PageRank scores are accessible through many different traversals, and hence are less reliant on backtracking walks than nodes given lower PageRank scores. We now show that this result carries over to the case where we allow teleporting, for a special class of graphs.

**Theorem 4** *Let  $W$  be the adjacency matrix of an undirected  $k$ -regular graph, with  $k \geq 2$ . For any  $\alpha \in [0, 1]$ , let us define the matrices  $\mathcal{P} = \alpha \mathcal{D}^{-1} \mathcal{B} + \frac{(1-\alpha)}{n} \mathbf{1} \mathbf{v}^T$  and  $P = \alpha D^{-1} W + \frac{(1-\alpha)}{n} \mathbf{1} \mathbf{1}^T$ , where  $\mathcal{D}$  is the diagonal matrix whose diagonal entries are the entries of the vector  $\mathcal{B} \mathbf{1}$ , and  $\mathbf{v} = L D^{-1} \mathbf{1}$ . Then, if we let  $\mathcal{P}^T \hat{\mathbf{y}} = \hat{\mathbf{y}}$  and  $P^T \mathbf{x} = \mathbf{x}$ , with  $\|\mathbf{x}\|_1 = \|\hat{\mathbf{y}}\|_1 = 1$ , it holds that*

$$L^T \hat{\mathbf{y}} = \mathbf{x}$$

where  $L$  is the source matrix associated with  $W$ .

Before proving the result, let us recall that in the case of undirected networks every undirected edge is considered as a pair of directed edges when building the matrices  $\mathcal{W}$  and  $\mathcal{B}$ .

**Proof** When  $\alpha = 1$  the result follows from [17]. If  $\alpha = 0$ , then from the description of  $P$  and  $\mathcal{P}$  it follows that  $\mathbf{x} = \frac{(1-\alpha)}{n} \mathbf{1}$ , and  $\hat{\mathbf{y}} = \frac{(1-\alpha)}{n} \mathbf{v}$ . The conclusion then follows from the definition of  $\mathbf{v}$  and the fact that  $D = L^T L$ .

Let us now consider the case of  $\alpha \in (0, 1)$ . The eigenproblems  $P^T \mathbf{x} = \mathbf{x}$  and  $\mathcal{P}^T \hat{\mathbf{y}} = \hat{\mathbf{y}}$  can be reformulated as two linear systems:

$$(I - \alpha W D^{-1}) \mathbf{x} = \frac{(1-\alpha)}{n} \mathbf{1}$$

and

$$(I - \alpha \mathcal{B}^T \mathcal{D}^{-1}) \hat{\mathbf{y}} = \frac{(1-\alpha)}{n} \mathbf{v},$$

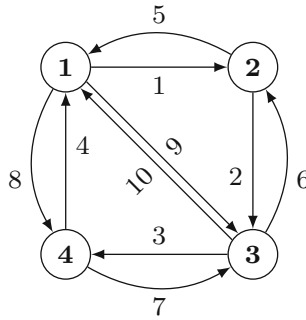


Fig. 1 Small example of undirected network for which  $L^T \hat{y} \neq \mathbf{x}$

where we have exploited the fact that  $W^T = W$ . Let us work on these individually. Concerning the first, we know that, since the graph is  $k$ -regular with  $k \geq 2$ , then  $W\mathbf{1} = k\mathbf{1}$  and consequently  $D^{-1} = k^{-1}I$ . Therefore, it is easily checked that

$$\mathbf{x} = \frac{(1 - \alpha)}{n} \sum_{r=0}^{\infty} \frac{\alpha^r}{k^r} W^r \mathbf{1} = n^{-1} \mathbf{1}.$$

Concerning the second linear system, we have that  $B\mathbf{1} = (k - 1)\mathbf{1}$  and  $D^{-1} = (k - 1)^{-1}I$ , since for every edge one can proceed using all the  $k$  edges adjacent to it, except for its reciprocal. In addition, we also have that  $B^T\mathbf{1} = (k - 1)\mathbf{1}$  due to the fact that the graph is  $k$ -regular. From the description of  $D$  above and the fact that  $L\mathbf{1} = \mathbf{1}$ , we have that  $\mathbf{v} = k^{-1}\mathbf{1}$ . Therefore,

$$\hat{y} = \frac{(1 - \alpha)}{kn} \sum_{r=0}^{\infty} \frac{\alpha^r}{(k - 1)^r} (B^T)^r \mathbf{1} = (kn)^{-1} \mathbf{1}.$$

The conclusion in Theorem 4 then follows from the fact that each node is the source of exactly  $k$  edges, and thus  $L^T \hat{y} = n^{-1} \mathbf{1} = \mathbf{x}$ . □

We now show with a small example that without the hypothesis of regularity the two steady states are no longer related in general when  $\alpha \in (0, 1)$ . For  $\alpha = 0, 1$  the conclusion still follows from an easy computation and from [17]. Let us consider the graph represented in Fig. 1, where we have represented each undirected edge as a pair of directed edges pointing in opposite directions. With this ordering of nodes and vertices, the source matrix reads

$$L^T = \left[ \begin{array}{cccc|cccc} 1 & & & & 0 & & 1 & & 1 & 0 & & \\ & 1 & & & 1 & 0 & & & 0 & 0 & & \\ & & 1 & & & 1 & 0 & & 0 & 1 & & \\ & & & 1 & & & 1 & 0 & 0 & 0 & & \end{array} \right].$$

From the symmetries in the graph it follows that, when working on the nodes, the entries of the PageRank vector  $\mathbf{x} = (x_i)$  are such that  $x_1 = x_3$  and  $x_2 = x_4$ ; exploiting those symmetries when working on the edges, it follows that the non-backtracking PageRank-like vector  $\hat{y} = (\hat{y}_i)$  will have  $\hat{y}_1 = \hat{y}_3 = \hat{y}_6 = \hat{y}_8$ ,  $\hat{y}_2 = \hat{y}_4 = \hat{y}_5 = \hat{y}_7$  and  $\hat{y}_9 = \hat{y}_{10}$ . Using these relationships, the linear systems of interest may be solved to give

$$x_1 = \frac{3(1 + \alpha)}{4(3 + 2\alpha)} \quad x_2 = \frac{3 + \alpha}{4(3 + 2\alpha)},$$

and

$$\hat{y}_1 = \frac{\alpha^2 + 2\alpha + 3}{12(\alpha^2 + 2\alpha + 2)}, \quad \hat{y}_2 = \frac{\alpha^2 + 3\alpha + 2}{12(\alpha^2 + 2\alpha + 2)},$$

$$\hat{y}_9 = \frac{\alpha^2 + \alpha + 1}{6(\alpha^2 + 2\alpha + 2)}.$$

The entries of  $\mathbf{y} := L^T \hat{\mathbf{y}}$  are then

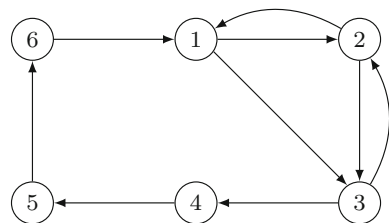
$$y_1 = y_3 = \frac{2\alpha^2 + 4\alpha + 3}{6(\alpha^2 + 2\alpha + 2)}, \quad y_2 = y_4 = \frac{\alpha^2 + 2\alpha + 3}{6(\alpha^2 + 2\alpha + 2)}.$$

It is easily verified that  $y_1 = x_1$  and  $y_2 = x_2$  if and only if  $\alpha = 0, 1$ ; hence the result in Theorem 4 does not hold for  $\alpha \in (0, 1)$ . We stress, however, that the ranking is preserved: nodes 1 and 3 rank first, while 2 and 4 rank second. The next example shows that this is not a universal behavior of the newly introduced measure, as we will also see experimentally in Sect. 6. Let us consider the network with  $n$  nodes whose adjacency matrix is

$$W = \left[ \begin{array}{ccc|cccc} 0 & 1 & 1 & & & & & & & & \\ 1 & 0 & 1 & & & & & & & & \\ 0 & 1 & 0 & 1 & & & & & & & \\ \hline & & & 0 & 1 & & & & & & \\ & & & & & \ddots & & & & & \\ & & & & & & \ddots & & & & \\ & & & & & & & \ddots & & & \\ & & & & & & & & 1 & & \\ 1 & & & & & & & & & 0 & \end{array} \right].$$

In Fig. 2 we illustrate the case  $n = 6$ . It can be proved analytically that when there is no teleporting and  $\alpha = 1$ , the standard PageRank vector  $\mathbf{x} = (x_i)$  is such that  $x_1 = x_2 = x_3 > x_4 = x_5 \cdots = x_n$ . This can be also be argued from first principles by considering the long-time frequency with which nodes are visited. After node 4 has been reached, a walk always continues with nodes 5, 6, ...,  $n$ , which explains why  $x_4 = x_5 = \cdots = x_n$ . A visit to node 1 is equally likely to be followed by a visit to nodes 2 or 3. A visit to node 2 is equally likely to be followed by a visit to nodes 1 or 3. A visit to node 3 is equally likely to be followed by a visit to node 2 or (after passing through 4, 5, ...,  $n$ ) node 1. Hence  $x_1 = x_2 = x_3$ , and this value exceeds the other frequencies. On the other hand, the new measure is such that  $y_1 = y_3 > y_2 = y_4 = y_5 = \cdots = y_n$ . Here, in the absence of backtracking the same argument explains why  $y_4 = y_5 = \cdots = y_n$ . The key difference now is that a visit to nodes 1 or 3 can no longer be followed by a visit to node 2 if node 2 was used on the previous step. So, in the long term, although node 2 receives more visits than nodes 4, 5, ...,  $n$  it loses out to nodes 1 and 3. Therefore the rankings derived from the two measures are different. The same

**Fig. 2** Small example of network for which the rankings induced by  $\mathbf{x}$  and  $\mathbf{y}$  are different for all  $\alpha \in (0, 1]$



conclusion holds when we take teleportation into account; in this situation similar arguments show that  $x_1 > x_2 = x_3 > x_4 > \dots > x_n$  and  $y_1 > y_3 > y_2 > y_4 > y_5 > \dots > y_n$ .

Before moving on to the numerical tests on real-world networks, we discuss in the next section how to implement the most expensive operation performed by an iterative solver when solving (8): matrix-vector multiplications involving  $\mathcal{B}^T$ .

### 5 On the Computation of $\mathcal{B}^T \mathbf{v}$

The Hashimoto matrix  $\mathcal{B}$  built from a directed network may be very large, especially if the original graph was large to begin with and contained several dangling nodes. Indeed, to remove such nodes one needs to add  $n$  edges to the network for each dangling node. While performing this correction to the graph in the node space corresponds to a low rank modification of the adjacency matrix  $A$ , it corresponds to adding  $n$  rows and columns to the non-backtracking edge-matrix for each of the dangling nodes originally present in the digraph.

In the following we describe how to label the edges in the network in order to highlight the structure and sparsity of  $\mathcal{B}$  and exploit such features in the computations. The notation adopted will be as follows:

- $n$  is the total number of nodes in the digraph;
- $k$  is the number of dangling nodes;
- $\tilde{m}$  is the number of edges *before* the correction of the dangling nodes, i.e., the number of edges in the graph associated with  $A$ ;
- $K \leq \tilde{m}$  is the number of edges in the graph represented by  $A$  that point to the dangling nodes.

We will assume that the non-dangling nodes are labelled as  $\{1, 2, \dots, n-k\}$  and the remaining  $k$  nodes  $\{n-k+1, \dots, n\}$  are the dangling nodes. Let us further assume that the  $\tilde{m} - K$  edges that do not point to the dangling nodes in the graph are labelled as  $\{1, 2, \dots, \tilde{m} - K\}$  so that the remaining  $K$  edges  $\{\tilde{m} - K + 1, \dots, \tilde{m}\}$  are those pointing to the dangling nodes. With this labelling, the matrix  $A$  will have the following structure:

$$A = \left[ \begin{array}{c|c} A_{11} & A_{12} \\ \hline 0_{k,n-k} & 0_{k,k} \end{array} \right]$$

where  $A_{11} \in \mathbb{R}^{(n-k) \times (n-k)}$  is the adjacency matrix of the graph obtained by removing the dangling nodes and the edges pointing to them and  $A_{12} \in \mathbb{R}^{(n-k) \times k}$  describes the edges pointing to the dangling nodes from the non-dangling nodes in the network. We then have that the source and target matrices of  $A$ , that we will denote by  $L(A)$  and  $R(A)$ , respectively, have the following structure:

$$L(A) = \left[ \begin{array}{c|c} L_1 & 0_{\tilde{m}-K,k} \\ \hline L_2 & 0_{K,k} \end{array} \right] \in \mathbb{R}^{\tilde{m} \times n}$$

and

$$R(A) = \left[ \begin{array}{c|c} R_1 & 0_{\tilde{m}-K,k} \\ \hline 0_{K,n-k} & R_2 \end{array} \right] \in \mathbb{R}^{\tilde{m} \times n},$$

where  $L_1, R_1 \in \mathbb{R}^{(\tilde{m}-K) \times (n-k)}$ ,  $L_2 \in \mathbb{R}^{K \times (n-k)}$ , and  $R_2 \in \mathbb{R}^{K \times k}$ .

After the correction of the dangling nodes, the adjacency matrix associated with the new graph  $G$  will be  $W = A + \chi \mathbf{1}^T$  with  $\chi = \sum_{i=n-k+1}^n \mathbf{e}_i$ , where  $\mathbf{e}_i \in \mathbb{R}^n$  is the  $i$ th vector of the standard basis of  $\mathbb{R}^n$ . The source and target matrices associated with  $W$  are

$$L = L(W) = \left[ \begin{array}{c} L(A) \\ \mathbf{1}_n \otimes \Delta^T \end{array} \right] \text{ and } R = R(W) = \left[ \begin{array}{c} R(A) \\ I_n \otimes \mathbf{1}_k \end{array} \right],$$

where

$$\Delta = [\mathbf{e}_{n-k+1} | \dots | \mathbf{e}_n] = \left[ \begin{array}{c} 0_{n-k,k} \\ I_k \end{array} \right] \in \mathbb{R}^{n \times k}.$$

With this ordering of the nodes and edges, the edge matrix  $\mathcal{W} = RL^T$  has the form

$$\mathcal{W} = \left[ \begin{array}{cccc} R_1 L_1^T & R_1 L_2^T & 0 & 0 \\ 0 & 0 & \mathbf{1}_{n-k}^T \otimes R_2 & \mathbf{1}_k^T \otimes R_2 \\ L_1^T \otimes \mathbf{1}_k & L_2^T \otimes \mathbf{1}_k & 0 & 0 \\ 0 & 0 & \mathbf{1}_{n-k}^T \otimes I_k \otimes \mathbf{1}_k & \mathbf{1}_k^T \otimes I_k \otimes \mathbf{1}_k \end{array} \right]$$

and thus the matrix  $\mathcal{B}^T$  involved in the computations is

$$\mathcal{B}^T = \left[ \begin{array}{cccc} L_1 R_1^T - S_1 & 0 & L_1 \otimes \mathbf{1}_k^T & 0 \\ L_1 R_2^T & 0 & (L_2 \otimes \mathbf{1}_k^T) - S_2 & 0 \\ 0 & (\mathbf{1}_{n-k} \otimes R_2^T) - S_2^T & 0 & \mathbf{1}_{n-k}^T \otimes I_k \otimes \mathbf{1}_k \\ 0 & \mathbf{1}_k \otimes R_2^T & 0 & (\mathbf{1}_k^T \otimes I_k \otimes \mathbf{1}_k) - S_4 \end{array} \right],$$

where the matrices  $S_1, S_2, S_4$ , which effect the removal of  $\mathcal{W} \circ \mathcal{W}^T$  from  $\mathcal{W} = \mathcal{B} + \mathcal{W} \circ \mathcal{W}^T$ , have the form

$$\begin{aligned} S_1 &= (L_1 R_1^T) \circ (L_1 R_1^T)^T, \\ S_2 &= (L_2 \otimes \mathbf{1}_k^T) \circ (\mathbf{1}_{n-k}^T \otimes R_2), \\ S_4 &= (\mathbf{1}_k^T \otimes I_k \otimes \mathbf{1}_k) \circ (\mathbf{1}_k^T \otimes I_k \otimes \mathbf{1}_k)^T. \end{aligned}$$

We note that  $S_1$  and  $S_4$  are square and symmetric, whilst  $S_2$  is rectangular, in general. All three matrices contain at most one element equal to one in each row. We show below that, because they are highly structured, we do not need to explicitly form the matrices  $S_2$  and  $S_4$  in order to compute their action on a vector.

Let us describe the matrices  $S_2 \in \mathbb{R}^{K \times k(n-k)}$  and  $S_4 \in \mathbb{R}^{k^2 \times k^2}$  in more detail. From the structure of  $(L_2 \otimes \mathbf{1}_k^T)$  and  $(\mathbf{1}_{n-k}^T \otimes R_2)$  it can be deduced that the ones in the matrix  $S_2$  are found in position  $(e, (i-1)k + j)$ , where  $e \in \{\tilde{m} - K + 1, \dots, \tilde{m}\}$  is the label of the edge  $i \rightarrow n - k + j$  that points from a certain node  $i \in \{1, 2, \dots, n - k\}$  to a dangling node  $(n - k + j)$  for  $j \in \{1, 2, \dots, k\}$ . Similarly, the matrix  $S_4 = (\mathbf{1}_k^T \otimes I_k \otimes \mathbf{1}_k) \circ (\mathbf{1}_k \otimes I_k \otimes \mathbf{1}_k^T)$  has a one in position  $((i-1)k + j, (j-1)k + i)$  for all  $i, j = 1, 2, \dots, k$ . Therefore, pre-multiplying a vector by these matrices corresponds to accessing certain entries of the vector.

In order to compute the non-backtracking PageRank vector, we can either solve a sparse linear system or use the power method to obtain the non-trivial eigenvector associated with the eigenvalue 1. In either case, we need to implement a matrix-vector product involving  $\mathcal{B}^T$  and a rescaled vector  $\mathbf{v} := \mathcal{D}^\dagger \bar{\mathbf{v}}$ . The rescaling can be performed at each iteration once the column sums of  $\mathcal{B}^T$  have been computed. Recall that we do not need to remove zero rows from  $\mathcal{B}$ , equivalently, columns from  $\mathcal{B}^T$ , as it is well known that the solution to the pseudo-PageRank problem (solve while retaining the zero columns in  $\mathcal{B}^T$ ) only differs from the

solution to the PageRank problem (solve after “correcting” the zero columns) by a positive multiplicative constant; see Theorem 1. Let  $\mathbf{v} = \mathcal{D}^\dagger \bar{\mathbf{v}} = [\mathbf{q}^T, \mathbf{r}^T, \mathbf{s}^T, \mathbf{t}^T]^T \geq \mathbf{0}$  be a non-zero vector with  $\mathbf{q} \in \mathbb{R}^{\tilde{m}-K}$ ,  $\mathbf{r} \in \mathbb{R}^K$ ,  $\mathbf{s} \in \mathbb{R}^{k(n-k)}$ , and  $\mathbf{t} \in \mathbb{R}^{k^2}$ . Then, at each step, we compute

$$\begin{aligned} \mathcal{B}^T \mathbf{v} &= \begin{bmatrix} L_1 R_1^T \mathbf{q} - S_1 \mathbf{q} + (L_1 \otimes \mathbf{1}_k^T) \mathbf{s} \\ L_2 R_1^T \mathbf{q} + (L_2 \otimes \mathbf{1}_k^T) \mathbf{s} - S_2 \mathbf{s} \\ (\mathbf{1}_{n-k} \otimes R_2^T) \mathbf{r} - S_2^T \mathbf{r} + (\mathbf{1}_{n-k} \otimes I_k \otimes \mathbf{1}_k^T) \mathbf{t} \\ (\mathbf{1}_k \otimes R_2^T) \mathbf{r} + (\mathbf{1}_k \otimes I_k \otimes \mathbf{1}_k^T) \mathbf{t} - S_4 \mathbf{t} \end{bmatrix} \\ &= \begin{bmatrix} L_1 (R_1^T \mathbf{q}) - S_1 \mathbf{q} + L_1 \widehat{\mathbf{s}} \\ L_2 (R_1^T \mathbf{q}) + L_2 \widehat{\mathbf{s}} - S_2 \mathbf{s} \\ \mathbf{1}_{n-k} \otimes (R_2^T \mathbf{r}) - S_2^T \mathbf{r} + \mathbf{1}_{n-k} \otimes \widehat{\mathbf{t}} \\ \mathbf{1}_k \otimes (R_2^T \mathbf{r}) + \mathbf{1}_k \otimes \widehat{\mathbf{t}} - S_4 \mathbf{t} \end{bmatrix} \end{aligned}$$

where  $(\widehat{\mathbf{s}})_i = \sum_{j=1}^k s_{(i-1)k+j}$  and  $(\widehat{\mathbf{t}})_i = \sum_{j=1}^k t_{(i-1)k+k}$ .

### 6 Numerical Experiments

We now present numerical results to compare the performance of non-backtracking PageRank in Definition 1 with standard PageRank. We perform our tests on four road networks corresponding to the cities of Birmingham (UK), Philadelphia (USA), Austin (USA), and the Federal State of Hesse (Germany). The data is available in [32]. Here, the edges represent roads and nodes represent road intersections (primal graph).

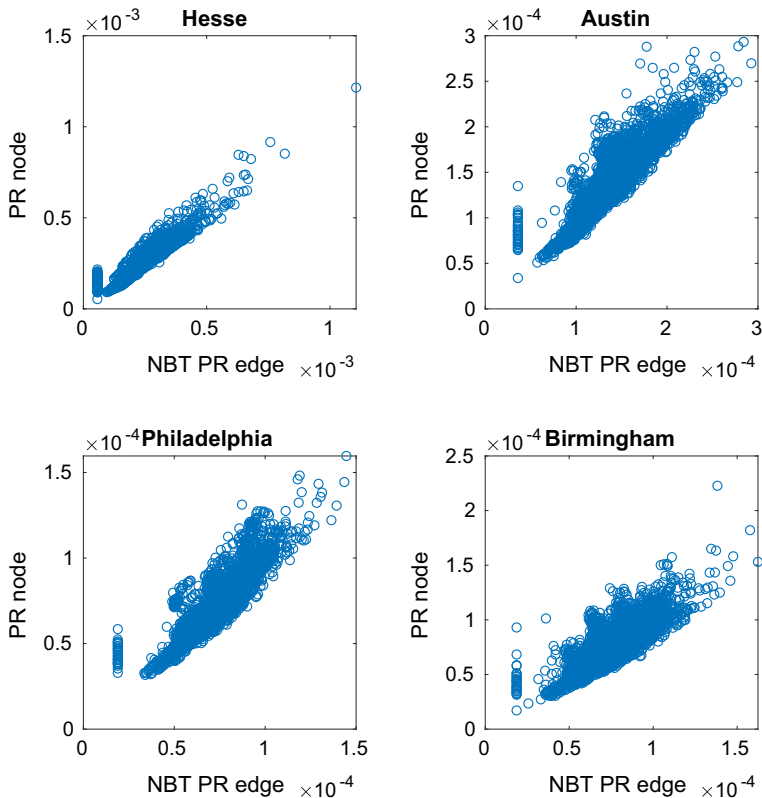
We computed both the new measure and standard PageRank as the solution of a linear system; see equations (5) and (8). For both problems we used the MATLAB implementation of the linear solver GMRES without restart and we gave as input a function handle that describes the matrix multiplication  $\mathcal{B}^T(\mathcal{D}^{-1}\mathbf{v})$  and  $W^T(\mathcal{D}^{-1}\mathbf{v})$  for vectors  $\mathbf{v}$  of appropriate size. Both these functions were implemented exploiting the structure of the matrices  $\mathcal{B}$  (see Sect. 5) and  $W = A + \chi \mathbf{1}^T$ .

We used the default setting of  $10^{-6}$  for the tolerance in the stopping criterion and allowed for a maximum of 100 iterations. We select  $\alpha = 0.75$  for the teleporting parameter. Properties of the networks used in the tests are summarized in Table 1, where we report the number of nodes,  $n$ , the number of dangling nodes,  $k$ , the number of source nodes,  $s = n - N$ , the number,  $\tilde{m}$ , of edges in the network represented by  $A$ , and the number,  $K$ , of edges pointing to the dangling nodes. Moreover, we report the number,  $M$ , of dangling edges in the network and the fraction of edges in the graph that have a reciprocal:  $\delta := \frac{\mathbf{1}^T(A \circ A^T)\mathbf{1}}{\tilde{m}}$ . So, for example, the value  $\delta = 0.19$  for the Federal State of Hesse indicates that 81% of the road connections are one-way only. All experiments were performed using MATLAB Version 9.4.0.813654 (R2018a) on an HP EliteDesk running Scientific Linux 7.5 (Nitrogen), a 3.2 GHz Intel Core i7 processor, and 4 GB of RAM.

In Fig. 3 we scatter plot the new measure (NBT PR edge), projected as in (9), against standard PageRank (PR node). The measures show a strong overall linear correlation; the Pearson’s correlation coefficients are  $\rho = 0.94$  for the Federal State of Hesse,  $\rho = 0.90$  for both Austin and Philadelphia, and  $\rho = 0.81$  for Birmingham. However, it is readily seen that the two measures do not completely agree on the rankings. Indeed, the intersection of the sets containing the IDs of the top ten ranked nodes according to the two measures

**Table 1** Description of the dataset

	$n$	$k$	$\tilde{m}$	$K$	$M$	$s$	$\delta$
FS of Hesse	4660	1	6674	1	245	1	0.19
Austin	7388	4	18,956	4	405	3	0.88
Philadelphia	13,389	0	40,003	0	178	0	0.94
Birmingham	14,639	0	33,937	0	1346	6	0.76

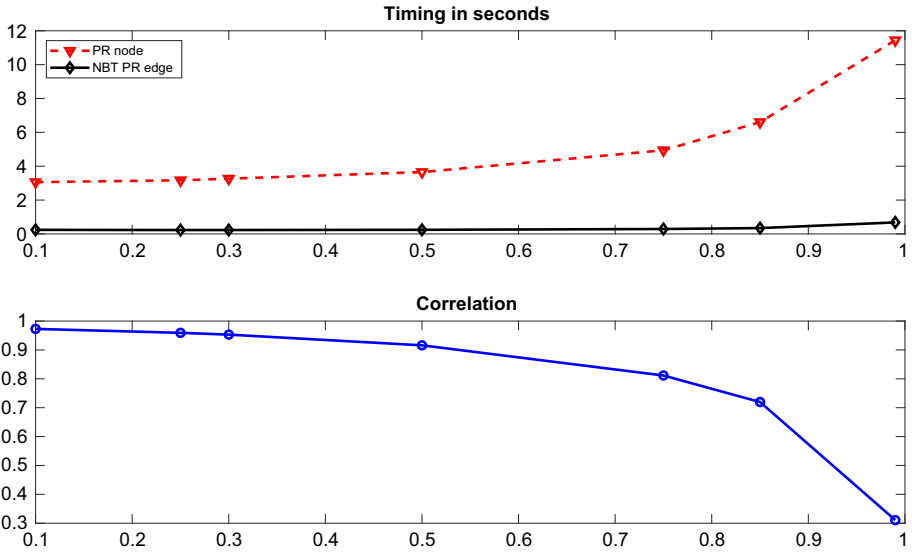
**Fig. 3** Scatter plot of the standard PageRank (PR node) vector  $\mathbf{x}$  in (5) versus the non-backtracking PageRank-like vector (NBT PR edge)  $\mathbf{y}$  in (9) both computed with  $\alpha = 0.75$ 

for Birmingham, the Federal State of Hesse, Austin and Philadelphia contain 8, 3, 5 and 6 elements, respectively. It is beyond the scope of this paper to determine which measure is returning the most meaningful results, since there is no ground truth for comparison. However, it may be argued that the non-backtracking version of the measure is the best fit for road networks since, as we mentioned in the introduction, the concept of non-backtracking walks is consistent with the typical behavior of a driver or a pedestrian traversing a city. It is notable that for all four scatter plots the low-ranked nodes are assigned the same score by non-backtracking PageRank, whilst they are being assigned different values by the classic version of the algorithm. A further analysis of the data showed that exactly  $M + s$  nodes fall in this class; these are the  $s$  source nodes and the  $M$  reciprocated leaves (i.e., nodes that have only one incoming link that is the reciprocal of their only outgoing edge) in the original graph associated with  $A$ .



**Table 2** Timings (in seconds) and number of GMRES iterations to compute the PageRank ranking vector defined at (5) and its non-backtracking version (8)

	PageRank		NBT PageRank		
	Time	Iter.	Time	Pre-proc.	Iter
FS of Hesse	6.39e-01	38	1.04e-01	6.85e-03	38
Austin	1.39	31	9.67e-02	9.07e-02	32
Philadelphia	4.71	28	1.27e-01	4.09e-01	30
Birmingham	5.18	29	6.46e-02	2.10e-01	31



**Fig. 4** Top: evolution of timings (in seconds) required to compute the vectors  $x$  and  $y$  for different choices of the damping parameters  $\alpha = 0.1, 0.25, 0.3, 0.5, 0.75, 0.85, 0.99$ . Bottom: evolution of Pearson’s correlation coefficient between  $x$  and  $y$  for  $\alpha = 0.1, 0.25, 0.3, 0.5, 0.75, 0.85, 0.99$

We now turn to computation time and the number of iterations required to achieve convergence of GMRES. Table 2 shows the results. We see that standard PageRank requires slightly fewer iterations than its non-backtracking counterpart. However, the total computation time for non-backtracking PageRank, including pre-processing time to build the matrices  $L_1, R_1, S_1, L_2$  and  $R_2$ , is significantly smaller than for standard PageRank. We attribute this surprising result to the fact that the sparsity and structure of the matrix  $B$  can be exploited by the iterative solver.

Finally, in Fig. 4 we display the evolution of the computation time required to achieve convergence of GMRES for the two algorithms as we let the teleportation parameter  $\alpha$  vary, as well as the evolution of the Pearson’s correlation coefficient between the two vectors for the city of Birmingham. Here we let  $\alpha = 0.1, 0.25, 0.3, 0.5, 0.75, 0.85, 0.99$ . Table 3 reports the number of iterations required to achieve convergence. Similar results were obtained for the other networks and are therefore omitted. We see from the upper plot of Fig. 4 that the newly introduced measure is computed more quickly than the measure we are comparing with, while Table 3 shows that the number of iterations required by GMRES to achieve convergence is comparable. Note that when  $\alpha = 0.99$ , in both cases GMRES iterated the maximum allowed number of times without achieving convergence. This follows from the

**Table 3** Number of iterations required to achieve convergence of GMRES

$\alpha$	0.1	0.25	0.3	0.5	0.75	0.85	0.99
PR node	5	8	9	15	29	45	–
NBT PR edge	6	9	10	16	31	47	–

fact that both systems become ill-conditioned as  $\alpha \rightarrow 1$ . The lower plot of Fig. 4 shows that the correlation between the two measures decreases as the value of  $\alpha$  increases. This may be explained by the fact that reducing the frequency of teleporting emphasizes the difference between the two underlying types of walk.

In summary, for the networks studied here, using the ideas from Sects. 4 and 5 we can incorporate non-backtracking into PageRank in order to obtain a distinct node ranking at reduced computational cost.

## 7 Conclusions

In this work we showed how to extend the definition of the PageRank algorithm to a non-backtracking framework by using the dual representation of a network and the Hashimoto matrix. We proved that the elimination of backtracking has no effect on the results in the case of undirected, regular networks, but also showed by example that non-backtracking can affect the ranking for the generic case of non-regular networks. We described explicitly how to exploit structure and sparsity in order to compute the new PageRank measure and showed through numerical experiments using real world transport networks and a GMRES solver that the computational cost is not increased. Future work will focus on (i) further tests on networks from a range of application fields, (ii) development of customized iterative methods for the structured linear systems that arise, (iii) non-backtracking *personalized PageRank* via a nonuniform teleportation distribution vector.

**Acknowledgements** The authors thank the anonymous referees for their valuable suggestions. We also thank Ernesto Estrada for pointing us to the datasets used in the numerical experiments. Francesca Arrigo thanks Mariarosa Mazza and Alessandro Buccini for helpful discussions.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Alon, N., Benjamini, I., Lubetzky, E., Sodin, S.: Non-backtracking random walks mix faster. *Commun. Contemp. Math.* **09**, 585–603 (2007). <https://doi.org/10.1142/S0219199707002551>
- Angel, O., Friedman, J., Hoory, S.: The non-backtracking spectrum of the universal cover of a graph. *Trans. Am. Math. Soc.* **326**, 4287–4318 (2015)
- Arrigo, F., Grindrod, P., Higham, D.J., Noferini, V.: Non-backtracking walk centrality for directed networks. *J. Complex Netw.* **6**(1), 54–78 (2018)
- Bowen, R., Lanford, O.E.: Zeta functions of restrictions of the shift transformation. In: Chern, S.S., Smale, S. (eds.) *Global Analysis: Proceedings of the Symposium in Pure Mathematics of the American Mathematical Society*, University of California, Berkeley, 1968, pp. 43–49. American Mathematical Society (1970)
- Constantine, P.G., Gleich, D.F.: Random alpha PageRank. *Internet Math.* **6**(2), 189–236 (2009)

6. Crisostomi, E., Kirkland, S., Shorten, R.: A Google-like model of road network dynamics and its application to regulation and control. *Int. J. Control* **84**(3), 633–651 (2011)
7. Crucitti, P., Latora, V., Porta, S.: Centrality in networks of urban streets. *Chaos Interdiscip. J. Nonlinear Sci.* **16**(1), 015113 (2006)
8. Del Corso, G.M., Gulli, A., Romani, F.: Fast PageRank computation via a sparse linear system. *Internet Math.* **2**(3), 251–273 (2005)
9. Gleich, D.F.: PageRank beyond the web. *SIAM Rev.* **57**(3), 321–363 (2015). <https://doi.org/10.1137/140976649>
10. Grindrod, P., Higham, D.J., Noferini, V.: The deformed graph Laplacian and its applications to network centrality analysis. *SIAM J. Matrix Anal. Appl.* **39**(1), 310–341 (2018)
11. Hashimoto, K.: Zeta functions of finite graphs and representations of p-adic groups. In: *Automorphic Forms and Geometry of Arithmetic Varieties*, pp. 211–280. Elsevier, Amsterdam (1989)
12. Horn, R.A., Johnson, C.R.: *Matrix Analysis*. Cambridge University Press, Cambridge (1985)
13. Horton, M.D.: Ihara zeta functions on digraphs. *Linear Algebra Appl.* **425**, 130–142 (2007)
14. Horton, M.D., Stark, H.M., Terras, A.A.: What are zeta functions of graphs and what are they good for? In: Bertolaiko, G., Carlson, R., Fulling, S.A., Kuchment, P. (eds.) *Quantum Graphs and Their Applications*, Contemporary Mathematics, vol. 415, pp. 173–190. American Mathematical Society, Providence, RI (2006)
15. Ipsen, I.C.F., Kirkland, S.: Convergence analysis of a PageRank updating algorithm by Langville and Meyer. *SIAM J. Matrix Anal. Appl.* **27**(4), 952–967 (2006)
16. Kawamoto, T.: Localized eigenvectors of the non-backtracking matrix. *J. Stat. Mech. Theory Exp.* **2016**, 023404 (2016)
17. Kempton, M.: Non-backtracking random walks and a weighted Ihara’s theorem. *Open J. Discrete Math.* **6**(04), 207–226 (2016)
18. Krzakala, F., Moore, C., Mossel, E., Neeman, J., Sly, A., Zdeborová, L., Zhang, P.: Spectral redemption: clustering sparse networks. *Proc. Natl. Acad. Sci.* **110**, 20935–20940 (2013)
19. Langville, A.N., Meyer, C.D.: Deeper inside PageRank. *Internet Math.* **1**, 335–380 (2004)
20. Martin, T., Zhang, X., Newman, M.E.J.: Localization and centrality in networks. *Phys. Rev. E* **90**, 052808 (2014)
21. Morone, F., Makse, H.A.: Influence maximization in complex networks through optimal percolation. *Nature* **524**, 65–68 (2015)
22. Page, L., Brin, S., Motwani, R., Winograd, T.: The PageRank citation ranking: bringing order to the web. Technical Report, Stanford University (1998)
23. Pastor-Satorras, R., Castellano, C.: Distinct types of eigenvector localization in networks. *Sci. Rep.* **6**, 18847 (2016)
24. Saade, A., Krzakala, F., Zdeborová, L.: Spectral clustering of graphs with the Bethe Hessian. In: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N.D., Weinberger, K.Q. (eds.) *Advances in Neural Information Processing Systems*, vol. 27, pp. 406–414. Curran Associates, Inc. (2014)
25. Schlote, A., Crisostomi, E., Kirkland, S., Shorten, R.: Traffic modelling framework for electric vehicles. *Int. J. Control* **85**(7), 880–897 (2012)
26. Smilansky, U.: Quantum chaos on discrete graphs. *J. Phys. A Math. Theor.* **40**, F621 (2007)
27. Sodin, S.: Random matrices, non-backtracking walks, and the orthogonal polynomials. *J. Math. Phys.* **48**, 123503 (2007)
28. Stark, H., Terras, A.: Zeta functions of finite graphs and coverings. *Adv. Math.* **121**(1), 124–165 (1996)
29. Tarfulea, A., Perlis, R.: An Ihara formula for partially directed graphs. *Linear Algebra Appl.* **431**, 73–85 (2009)
30. Terras, A.: *Harmonic Analysis on Symmetric Spaces—Euclidean Space, the Sphere, and the Poincaré Upper Half-Plane*, 2nd edn. Springer, New York (2013)
31. Torres, L., Suarez-Serrato, P., Eliassi-Rad, T.: Graph distance from the topological view of non-backtracking cycles. [arXiv:1807.09592](https://arxiv.org/abs/1807.09592) [cs.SI] (2018)
32. Transportation Networks for Research Core Team: Transportation Networks for Research. <https://github.com/bstabler/TransportationNetworks>. Accessed 25 July 2018
33. Watanabe, Y., Fukumizu, K.: Graph zeta function in the Bethe free energy and loopy belief propagation. In: Bengio, Y., Schuurmans, D., Lafferty, J., Williams, C., Culotta, A. (eds.) *Advances in Neural Information Processing Systems*, vol. 22, pp. 2017–2025. Curran Associates, Inc. (2009)
34. Wills, R.S., Ipsen, I.C.F.: Ordinal ranking for Google’s PageRank. *SIAM J. Matrix Anal. Appl.* **30**(4), 1677–1696 (2009)