Journal of Applied and Computational Topology
https://doi.org/10.1007/s41468-023-00129-6

A flexible PageRank-based graph embedding framework closely related to spectral eigenvector embeddings

Disha Shur[1] · **Yufan Huang**[1] · **David F. Gleich**[1]

Received: 30 June 2022 / Revised: 16 January 2023 / Accepted: 7 June 2023

Abstract
We study a simple embedding technique based on a matrix of personalized PageRank vectors seeded on a random set of nodes. We show that the embedding produced by the leading singular vectors of an element-wise logarithm of this matrix is related to the spectral embedding of Laplacian eigenvectors for degree regular graphs. Moreover, this log-PageRank embedding procedure produces useful results for global graph visualization even when the spectral embedding does not. Most importantly, the general nature of this embedding strategy opens up many emerging applications, where eigenvector and spectral techniques may not be well established, to the PageRank-based relatives. For instance, similar techniques can be used on PageRank vectors from hypergraphs to get "spectral-like" embeddings.

Keywords Graphs · Networks · Spectral embedding · Laplacian eigenvectors · Personalized PageRank · Low dimensional embeddings

Mathematics Subject Classification 62M15 · 05C65 · 05C82 · 05C81 · 05C50 · 65F15

1 Introduction

The eigenvectors of the graph Laplacian are among the most widely used algorithmic measures of a graph. They are used to find cuts and clusters in a variety of settings (Shi

Disha Shur and Yufan Huang have contributed equally to this work.

✉ Disha Shur
dshur@purdue.edu

✉ David F. Gleich
dgleich@purdue.edu

Yufan Huang
huan1754@purdue.edu

[1] Computer Science Department, Purdue University, 305 N. University St., West Lafayette 47906, IN, USA

Published online: 21 July 2023

and Malik 2000; Chung 1992; Pothen et al. 1990). They give a signal basis for a graph (Hammond et al. 2011; Donnat et al. 2018). And one of their original uses was to draw informative pictures of graphs in a low dimensional space (Hall 1970; Koren 2003). These are all related to the idea of embedding the graph into a low dimensional space and recent uses have closely studied this embedding framework.

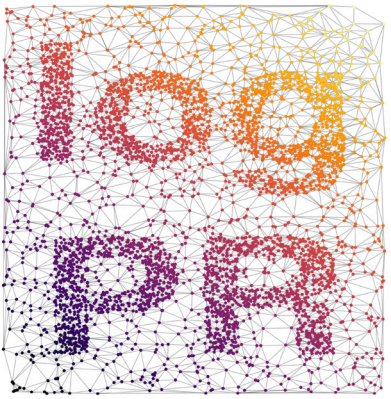
Likewise, PageRank is itself a widely used algorithmic measure on a graph (Brin and Page 1998). The uses are extremely diverse (Gleich 2015). Relationships between PageRank and spectral clustering are also known (Andersen et al. 2006; Mahoney et al. 2012; Gleich and Mahoney 2014). These exist because both techniques can be related to random walks, and seeded PageRank is a localized type of random walk, or random walk with restart (Tong et al. 2006).

In this manuscript, we study a particular type of relationship between a matrix of seeded PageRank vectors and the eigenvectors of the Laplacian matrix. Our log-PageRank embedding uses the singular vectors of the elementwise log of a random collection of seeded PageRank vectors. An example is in Fig. 1, which shows that log-PageRank embeddings resemble spectral embedding. Our manuscript establishes that this relationship is expected for degree-regular graphs (Sect. 5), which builds on a simpler characterization of log-PageRank values for the chain graph (Sect. 4).

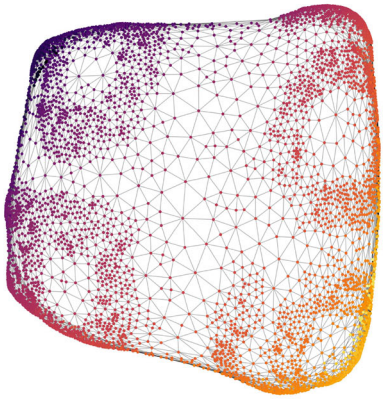
Taking the log of PageRank values has long been a standard practice. When Google published PageRank scores for websites, they were understood to represent an approximation of the log of Google's internal metrics (Bar-Yossef and Mashiach 2008). When PageRank was used in spam analysis, log scaling was used by Becchetti et al. (2008, Section 6.4). So a log-scaled analysis is not surprising. More recently, log-PageRank emerged as a topic in graph representation learning due to a relation between Skip-Gram embeddings and log-PageRank values (Zhou et al. 2017) and further efforts utilize singular value decompositions (Tsitsulin et al. 2021) as we do. We discuss these relationships in more details in Sect. 8.

Analyzing the singular vectors of personalized PageRank vectors under log-scaling presents interesting challenges from a technical perspective. We are able to study these by leveraging recent tools in random matrix theory (Tropp 2012; Drineas and Ipsen 2019). The main result shows that for degree regular graphs, as α gets closer to 1, the log-PageRank embeddings resemble the spectral embeddings for an appropriate number of samples, which is much lesser than the number of nodes.

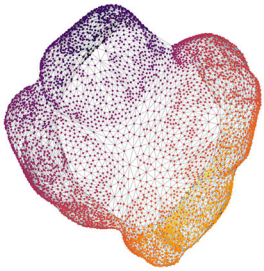
These log-PageRank embeddings offer a different set of computational tradeoffs compared with eigenvectors. In Fig. 2 we present a network where embedding using spectral techniques fails to give any significant information in a visualization but the log-PageRank embeddings give a useful global picture of the graph. In Fig. 3, spectral embeddings of the US road network show some structure, and the log-PageRank embeddings show arguably greater structure. This serves as a motivation to deploy this procedure where spectral embeddings fail and yet an explainable visualization of the network is required. Our log-PageRank embeddings are furthermore easy to specialize in new ways. Indeed, a closely related methodology to these log-PageRank embeddings was previously used in Fountoulakis et al. (2020) to compare spectral clustering with alternatives. For instance, it is easy to study a variety of localized log-PageRank embeddings that are only seeded in a specific region of the graph. These will pull in other nearby regions as suggested by the PageRank vectors instead of more



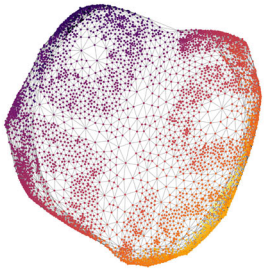
(a) A planar graph with 5000 nodes and 14962 edges with the words "log PR."



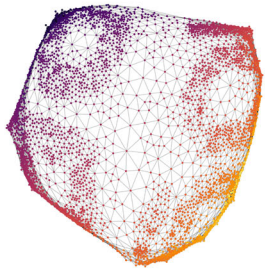
(b) The spectral embedding of the graph from the Laplacian eigenvectors.



(c) The log-PageRank embedding for $\alpha = 0.85$.



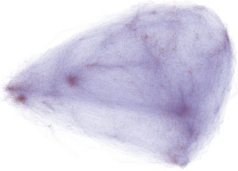
(d) The log-PageRank embedding for $\alpha = 0.99$.



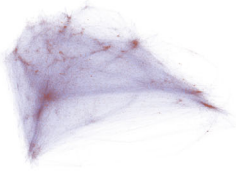
(e) The log-PageRank embedding for $\alpha = 0.999$.

Fig. 1 The embedding pictures have all the nodes colored with the same values to show relative position. The log-PageRank embedding uses singular values of the element-wise logarithm of seeded PageRank vectors. Our paper argues that the similarity between the embeddings shown in (**b**), (**d**) and (**e**) for the graph in (**a**) is expected through an approximation analysis The result in (**c**) is an easier to compute variation. The advantage the log-PageRank embeddings is that they can be deployed in many emerging data scenarios where spectral embeddings and eigenvectors are not as well established or may be computationally expensive but where analogues of random walks or PageRank may be possible, as in hypergraphs (see Fig. 12b)

(a) Spectral embedding



(b) log-PageRank embedding with $\alpha = 0.9$



(c) log-PageRank embedding with $\alpha = 0.99$

Fig. 2 The spectral embedding of an artist similarity network in (**a**) is overly-localized as often happens in real-world networks (Lang 2005) and results in a useless visualization. In comparison the log-PageRank embeddings with α not too close to 1, as shown in (**b**) and (**c**), show the global structure of the graph

(a) Spectral embedding (b) log-PageRank embedding with $\alpha = 0.9$ (c) log-PageRank embedding with $\alpha = 0.99$

Fig. 3 The US highway network has a useful spectral embedding in (**a**), but highly compresses important regions on the east and west coast. The left filament contain all roads in California, Oregon, and Washington and the right filament contains the entire northeast (New York, Boston, etc.), see Fountoulakis et al. (2020) for a labeled picture. In comparison, the log-PageRank embeddings spread these regions out for α between 0.85 and 0.99 as shown in (**b**) and (**c**)

brittle Dirichlet eigenvector approximations (Chung et al. 2011). As an example of this flexibility, we briefly explore using log-PageRank embeddings on hypergraphs to visualize their structure (Sect. 7).

In summary, the contributions and remainder of this paper discuss:

- the log-PageRank embedding framework (Sect. 3)
- a study of log-PageRank values on a chain graph that shows how log-PageRank values are related to graph distance (Sect. 4)
- an approximation analysis between log-PageRank embeddings and spectral embedding on d-regular graphs (Sect. 5)
- a computational study of similarities and differences between spectral and log-PageRank embeddings (Sect. 6)
- examples of log-PageRank embeddings in hypergraphs using hypergraph PageRank (Liu et al. 2021) (Sect. 7).

2 Preliminaries

In this manuscript we consider a connected, weighted or unweighted, undirected graph $G = (V, E)$ and its various spectral properties. Many of the matrices we will use are detailed in Table 1. We use subscripts to index entries of a matrix or a vector: let $\boldsymbol{A}_i$ denote the ith column of matrix $\boldsymbol{A}$, $\boldsymbol{A}_{ij}$ denote the (i, j)th entry of matrix $\boldsymbol{A}$. All norms of vectors and matrices without specific annotations are 2-norms.

PageRank The classical PageRank problem is defined as follows

Definition 1 (*see for example* Gleich (2015)) Let $\boldsymbol{P}$ be a column-stochastic matrix and $\boldsymbol{v}$ be a column-stochastic vector, then the PageRank problem is to find the solution $\boldsymbol{x}$ to the linear system

$$(\boldsymbol{I} - \alpha \boldsymbol{P})\boldsymbol{x} = (1 - \alpha)\boldsymbol{v} \tag{1}$$

where the solution $\boldsymbol{x}$ is called the PageRank vector, $\alpha \in (0, 1)$ is the teleportation parameter and $\boldsymbol{v}$ is the teleportation distribution vector.

Table 1 Notations

Notation	Description
$G = (V, E)$	Graph G with vertex set V and edge set E
n	number of vertices
$\boldsymbol{A}$	Adjacency matrix
$\boldsymbol{D}$	Diagonal degree matrix
$\boldsymbol{L}$	Laplacian, $\boldsymbol{D} - \boldsymbol{A}$
$\boldsymbol{W}$	Lazy random walk matrix $(1/2)(\boldsymbol{I} + \boldsymbol{A}\boldsymbol{D}^{-1})$
$\boldsymbol{P}$	Column stochastic transition matrix on G
$\boldsymbol{\pi}$	Solution to $\boldsymbol{P}\boldsymbol{\pi} = \boldsymbol{\pi}$ such that $\boldsymbol{\pi}$ is non-negative and sums to 1
$\boldsymbol{A}_{i:j}$	Matrix made up of columns $\boldsymbol{A}_i, \cdots, \boldsymbol{A}_j$
$\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n$	Columns of the identity matrix, n standard basis vectors of $\mathbf{R}^n$
$\boldsymbol{e}$	All-ones vector
$\boldsymbol{e}_u$	Indicator vector for node u
$\boldsymbol{x}(u, \alpha)$	Solution to $(\boldsymbol{I} - \alpha\boldsymbol{P})\boldsymbol{x} = (1 - \alpha)\boldsymbol{e}_u$
log .	Element-wise logarithm operator

By the definition above and the fact that all eigenvalues of a column-stochastic matrix have magnitude at most 1, $\boldsymbol{I} - \alpha\boldsymbol{P}$ is non-singular and the PageRank vector can be written as $\boldsymbol{x} = (1 - \alpha)(\boldsymbol{I} - \alpha\boldsymbol{P})^{-1}\boldsymbol{v}$. When the teleportation distribution $\boldsymbol{v}$ has support size 1, the PageRank problem is also called seeded PageRank or personalized PageRank and the corresponding solution $\boldsymbol{x}$ is a seeded PageRank vector or a personalized PageRank vector. For convenience, let $\boldsymbol{X}(\alpha)$ denote $(1 - \alpha)(\boldsymbol{I} - \alpha\boldsymbol{P})^{-1}$, $\boldsymbol{x}(u, \alpha)$ denote the personalized PageRank vector seeded on vertex u, in other words $\boldsymbol{x}(u, \alpha) = \boldsymbol{X}(\alpha)\boldsymbol{e}_u = (1 - \alpha)(\boldsymbol{I} - \alpha\boldsymbol{P})^{-1}\boldsymbol{e}_u$.

3 Log-PageRank embedding

Our study of log-PageRank embeddings uses the procedure detailed in Algorithm 1. It takes as input the graph $G = (V, E)$ and outputs the k-dimensional node embeddings. We randomly sample nodes of the graph, compute personalized PageRank vectors, and then compute an elementwise log of the resulting vectors. Then we compute an SVD of the entire set of sampled vectors. The non-dominant vectors give us our log-PageRank embedding. Note that a personalized PageRank vector has mathematically non-negative entries for a connected graph, so computing the log is always mathematically well defined. However, numerically, some of the elements may be sufficiently close to zero to cause an algorithm to return a floating point zero. For this reason, we often replace any zero entries with a value smaller than the smallest non-zero element returned before taking the log. This only occurs for small values of α and tends not to happen once α is close enough to one.

Note that our log-PageRank based technique offers freedom in the algorithm being used for calculation of PageRank vector. For instance, in Sect. 7 we will use a hyper-

Algorithm 1 Log-PageRank Embedding

Input: Graph adjacency matrix $\boldsymbol{A}$, Dimension of embedding k, Number of samples $s \geq k+1$ (we suggest $s = (10+k)\log n$), Teleportation parameter α
Output: Graph embedding $\boldsymbol{Z} \in \mathbb{R}^{n \times k}$

for $i = 1 \rightarrow s$ **do**
  $u \leftarrow$ `random sample of 1 to` n
  $\boldsymbol{X}_i \leftarrow$ `pagerank on` $\boldsymbol{A}$ `with seed` u`, teleportation param` α
  ▷ We use a single sparse LU on $\boldsymbol{I} - \alpha \boldsymbol{P}$ to compute PageRank
end for
$\boldsymbol{Y} \leftarrow \log.(\mathbf{X})$ ▷ Apply element-wise log on $\boldsymbol{X}$
$\boldsymbol{U}, \boldsymbol{\Sigma}, \boldsymbol{V} \leftarrow$ SVD of $\boldsymbol{Y}$
$\boldsymbol{Z} \leftarrow \boldsymbol{U}_{2:k+1}$
return $\boldsymbol{Z}$ ▷ Return left singular vectors of $\boldsymbol{Y}$

graph PageRank vector instead. Instead, it could also use seeded PageRank vectors only from a region of interest within the graph.

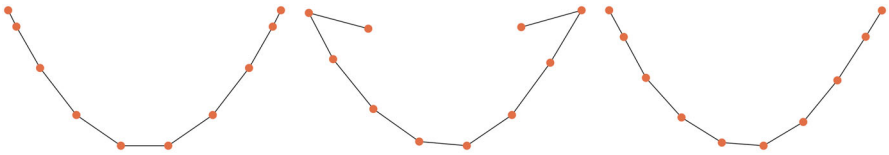
Parameters. The parameters in this technique are the dimension of embedding, k, the teleportation parameter, α, and the number of samples s. The dimension is entirely at a user's discretion. For the number of samples, we suggest scaling the $\log(n)$ term in Theorem 1 with the dimension of embedding, in other words taking at least $k \log(n)$ samples. For the teleportation parameter, we suggest use $\alpha \geq 0.9$, such as $\alpha = 0.99$ or $\alpha = 0.999$. Because we use many PageRank computations with large values of α, we find it pragmatic to compute a single sparse LU decomposition of the matrix $\boldsymbol{I} - \alpha \boldsymbol{P}$ to repeatedly solve systems. Apart from the PageRank computation, the runtime depends on the SVD of the PageRank matrix, for which any type of randomized SVD computation could be used to make it more efficient.

Intuition and analysis. The origin of this algorithm is from Fountoulakis et al. (2020), where Gleich and co-authors used the linearity of PageRank and the relationship with an expectation to study spectral-like embeddings of nonlinear operators using a similar process.

The idea behind the algorithm is that the matrix of samples should have substantial information from other eigenspaces beyond the dominant one and the SVD will return this information. Our study of this algorithm revealed that the log is essential to getting qualitatively *similar* pictures such as those in Fig. 1. We show in Sect. 5 that as α approaches 1, the log-PageRank embedding approximates the eigenvectors of the lazy random walk matrix $\boldsymbol{W}$. We illustrate a simple example that motivates a relationship between log-PageRank values and a notion of distance.

4 Log-PageRank on the chain graph

The chain graph is an extremely simple graph. Imagine vertices laid out on a line and connect each vertex to the two left and right vertices. At the ends of the line, the vertices only have degree one. See Fig. 4a. We developed a closed form expression for personalized PageRank on the chain graph and observed a linear dependence between the element-wise log of PageRank and the graph distance.



(a) Chain graph on 10 nodes showing coordinates generated by spectral embedding

(b) Chain graph embeddings generated by PageRank values (without log) at $\alpha = 0.99$

(c) Chain graph embeddings generated by log PageRank values at $\alpha = 0.99$

Fig. 4 The structure of Chain graph allows us to probe theoretically over the relation induced by the logarithm operation on PageRank. Notice that in (**c**), although a small graph, log of PageRank develops smoother embedding that are more similar to spectral embedding, as in (**a**), as compared to only PageRank embeddings (**b**)

For a chain graph of size $n > 2$, the linear system defined in Eq. (1) forms a second degree recurrence relation of the following form

$$\begin{cases} \frac{\alpha}{2}(\boldsymbol{x}_{i+1} + \boldsymbol{x}_{i-1}) = \boldsymbol{x}_i, \forall i \in [n] \setminus \{u\} \\ \frac{\alpha}{2}(\boldsymbol{x}_{u+1} + \boldsymbol{x}_{u-1}) = \boldsymbol{x}_u - (1-\alpha) \\ \sum_{i=1}^{n} \boldsymbol{x}_i = 1 \\ \boldsymbol{x}_0 = 0, \boldsymbol{x}_{n+1} = 0. \end{cases}$$

Solving the above gives the following closed form expression in terms of u, n, α.

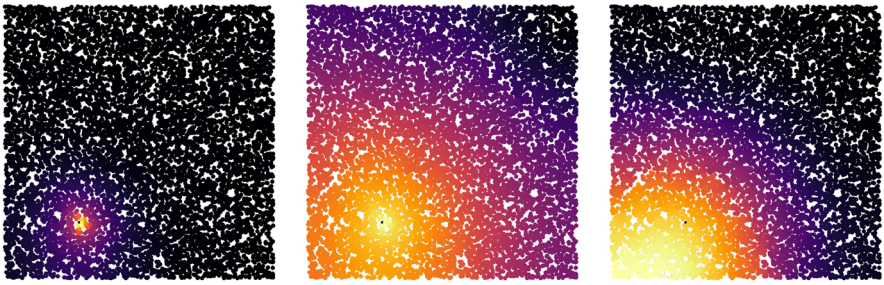
$$\boldsymbol{x}_i = \begin{cases} cf(i), i \in \{2, \ldots, u-1\} \\ \frac{c\alpha}{2}(f(u-1) + g(u+1)), i = u \\ cg(i), i \in \{u+1, \ldots, n-1\} \end{cases} \tag{2}$$

where

$$f(i) = \frac{(+)^{i-1} + (-)^{i-1}}{(+)^{u-1} + (-)^{u-1}}, g(i) = \frac{(+)^{n-i} + (-)^{n-i}}{(+)^{n-u} + (-)^{n-u}},$$
$$c = \sqrt{\frac{1-\alpha}{1+\alpha}}, (+) = \frac{1+\sqrt{1-\alpha^2}}{\alpha}, (-) = \frac{1-\sqrt{1-\alpha^2}}{\alpha}.$$

Notice that when α is far from 1, $(-) \approx 0$ for high powers, and when α is close to 1, $(+) \approx (-)$. In both ways we get the same approximation that

$$\boldsymbol{x}_i \approx \begin{cases} c(+)^{-|i-u|}, i \in [n] \setminus \{u\} \\ c\frac{\alpha}{(+)}, i = u \end{cases}.$$



(a) PageRank values for a 10000 node graph with 6 nearest neighbour with $\alpha = 0.999$

(b) Log of PageRank values for a 10000 node graph with 6 nearest neighbour with $\alpha = 0.999$

(c) Normalized Adjacency powers for the seed used in PageRank above and $p = 2000$

Fig. 5 Distance effect created by log of PageRank in a geometric graph. The normalized adjacency matrix power is $(\boldsymbol{D}^{-1/2}\boldsymbol{A}\boldsymbol{D}^{-1/2})^p\boldsymbol{v}$ where $\boldsymbol{v}$ is the same as the PageRank seed. Note the stronger similarity of (**b**, **c**). Also note the difference is at the boundary. The boundary is where we tend to see the biggest differences between log-PageRank and spectral embeddings

Then the logarithm of PageRank expression for $\boldsymbol{x}_i$ can be written as

$$\log \boldsymbol{x}_i \approx -|u - i| \log((+)) + \log(\sqrt{\frac{1-\alpha}{1+\alpha}}).$$

The above formulation indicates a linear relation between the log-PageRank and the distance from the seed node. This hints at log-PageRank being a good measure of the structure of the network around the seed node.

We quickly verify that log-PageRank resembles the notion of "distance" in a geometric graph. The graph is created by randomly sampling points and connecting every point to its 6 nearest neighbors. The difference between PageRank and log-PageRank in this context is illustrated in Fig. 5.

5 Relation between Log-PageRank embedding and spectral embedding

In this section, we theoretically illustrate the relation between log-PageRank Embedding and Spectral Embedding on a special class of graphs, d-regular graphs.

Recall that the lazy random walk is denoted by $\boldsymbol{W} = \frac{\boldsymbol{I}+\boldsymbol{A}\boldsymbol{D}^{-1}}{2}$. Our use of the lazy walk matrix is due to the simplicity in analyzing powers of the matrix because it is fundamentally aperiodic. A more intricate analysis would likely be able to remove the aperiodicity.

Let the transition probability matrix $\boldsymbol{P}$ of PageRank be the matrix $\boldsymbol{W}$, so that we analyze the PageRank vectors $(1-\alpha)(\boldsymbol{I} - \alpha\boldsymbol{W})^{-1}\boldsymbol{e}_u$. By a variety of existing analyses (Serra-Capizzano 2005; Gleich 2009), we know that $\lim_{\alpha\to 1^-} \boldsymbol{x}(u,\alpha) =$

$\lim_{\alpha\to 1^-}(1-\alpha)(\boldsymbol{I}-\alpha\boldsymbol{W})^{-1}\boldsymbol{e}_u = \boldsymbol{\pi}$. This extends to log by continuity. Thus, $\lim_{\alpha\to 1^-}\log.(\boldsymbol{x}(u,\alpha)) = \log.(\boldsymbol{\pi})$.

We continue our study on d-regular graphs. Because for d-regular graphs, $\boldsymbol{A}$ shares the same eigenvectors with $\boldsymbol{W}$, instead of analyzing the eigenvectors of $\boldsymbol{A}$ as spectral embedding does, we analyze the eigenvectors of $\boldsymbol{W}$ and connect them with log-PageRank Embedding.

In the following discussion, we define

- N is the number of randomly sampled log-PageRank vectors,
- $\boldsymbol{R}$ as the matrix formed by log PageRank vectors that are the result of algorithm 1 before the SVD and scaled to have unit 2-norm columns,
- $\boldsymbol{M} = \frac{1}{N}\boldsymbol{R}\boldsymbol{R}^T$, that is, a scaled Gram matrix of $\boldsymbol{R}$,
- $\boldsymbol{C} = \boldsymbol{W}^k{\boldsymbol{W}^k}^T$ and
- $\hat{\boldsymbol{C}} = \frac{n}{N}\boldsymbol{W}^k\boldsymbol{S}\boldsymbol{S}^T{\boldsymbol{W}^k}^T$ where $\boldsymbol{S}$ is a column sampling matrix with N columns and each column of $\boldsymbol{S}$ is a random unit basis vector $\boldsymbol{e}_i$.

Our theoretical proof can be broken down into three steps.

1. First we argue that $\hat{\boldsymbol{C}}$ approximates $\boldsymbol{C}$ when the number of sampled columns are not too small.
2. Second we show that $\boldsymbol{M}$ approximates $\hat{\boldsymbol{C}}$ when α is close to 1 and k is reasonably large.
3. Third we provide an upper bound on the angles between the subspace spanned by k dominant left singular vectors of $\boldsymbol{R}$ and k dominant eigenvectors of $\boldsymbol{W}$.

Note that $\hat{\boldsymbol{C}} = \frac{n}{N}\boldsymbol{B}\boldsymbol{B}^T$ where

$$\boldsymbol{B} = \left[\boldsymbol{W}^k\boldsymbol{e}_{i_1}\ \boldsymbol{W}^k\boldsymbol{e}_{i_2}\ \boldsymbol{W}^k\boldsymbol{e}_{i_3}\ \ldots\ \boldsymbol{W}^k\boldsymbol{e}_{i_N}\right]$$

and $i_1,\ldots,i_N$ are the indices of N randomly sampled columns. We first prove the following result that characterizes the approximation error between $\boldsymbol{C}$ and $\hat{\boldsymbol{C}}$. This proof is inspired by Constantine and Gleich (2014).

We note that the precise quantification of this bound should be used as guidance to help understand the nature of the relationships among the quantities involved, rather than a precise quantification.

Theorem 1 *Let $\varepsilon > 0, L \in (0,1)$ and $k = \Omega(\log(\frac{L}{n^{3/2}})/\log(\lambda_2))$ where λ_2 is the second largest eigenvalue of $\boldsymbol{W}$. Define the variance*

$$\nu^2 = \left\| \frac{1}{n}\sum_{i=1}^{n}\left(n(\boldsymbol{W}^k\boldsymbol{e}_i)(\boldsymbol{W}^k\boldsymbol{e}_i)^T - \boldsymbol{C}\right)^2 \right\|,$$

assume $\nu^2 > 0$ and let

$$\delta = \max(\frac{\nu^2}{\varepsilon}, (1+L)^2),$$

then using N samples where

$$N = \Omega(\frac{\delta}{\varepsilon}\log(2n))$$

implies that $\|\boldsymbol{C} - \hat{\boldsymbol{C}}\| \leq \varepsilon\|\boldsymbol{C}\|$ with high probability in the number of samples N.

Before proving this result, we first introduce a few useful lemmas. The first one is to bound the magnitude of columns of $\boldsymbol{W}^k$. Note that as $k \to \infty$, then $\boldsymbol{W}^k\boldsymbol{e}_i \to \boldsymbol{e}/n$ so $\|\boldsymbol{W}^k\boldsymbol{e}_i\| \to \frac{1}{\sqrt{n}}$. This lemma more precisely quantifies this convergence.

Lemma 1 *For any $L \in (0, 1)$, $k = \Omega(\log(\frac{L}{n^{3/2}})/\log(\lambda_2))$, and $\forall i \in [n]$, we have*

$$\|\boldsymbol{W}^k\boldsymbol{e}_i\| \leq (1 + L)\frac{1}{\sqrt{n}},$$

where λ_2 is the second largest eigenvalue of $\boldsymbol{W}$.

Proof Consider the eigenvalue decomposition $\boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T$ of $\boldsymbol{W}$, by Perron–Frobenius theory (Perron 1907; Frobenius 1912), since G is connected and $\boldsymbol{W}$ models a walk with self-loop, we have that $1 = \lambda_1 > |\lambda_i| \forall i > 1$, and $\boldsymbol{Q}_1 = \frac{\boldsymbol{e}}{\sqrt{n}}$. We observe that

$$\boldsymbol{W}^k\boldsymbol{e}_i = \boldsymbol{Q}\boldsymbol{\Lambda}^k\boldsymbol{Q}^T\boldsymbol{e}_i = \boldsymbol{Q}_1\boldsymbol{Q}_{i1} + \sum_{j=2}^{n}\lambda_j^k\boldsymbol{Q}_j\boldsymbol{Q}_{ij}.$$

Thus

$$\begin{aligned}\|\boldsymbol{W}^k\boldsymbol{e}_i\| &\leq \|\boldsymbol{Q}_1\boldsymbol{Q}_{i1}\| + \sum_{j=2}^{n}\lambda_j^k\|\boldsymbol{Q}_{ij}\boldsymbol{Q}_j\| \leq \frac{1}{\sqrt{n}} + \sum_{j=2}^{n}\lambda_j^k \\ &\leq \frac{1}{\sqrt{n}} + n\lambda_2^k \leq (1 + L)\frac{1}{\sqrt{n}}\end{aligned}$$

where the last inequality follows from our choice of k. □

Notice that to bound 2-norm of $\boldsymbol{C} - \hat{\boldsymbol{C}}$, we only need to bound $\lambda_{\max}(\boldsymbol{C} - \hat{\boldsymbol{C}})$ and $\lambda_{\max}(\hat{\boldsymbol{C}} - \boldsymbol{C})$, we introduce a useful matrix concentration inequality.

Theorem 2 (Matrix Bernstein: bounded case, Theorem 6.1 of Tropp (2012)) *Consider a finite sequence $\{\boldsymbol{X}_j\}$ of independent, random, self-adjoint matrices with dimension n. Assume that*

$$\mathbb{E}[\boldsymbol{X}_j] = 0 \text{ and } \lambda_{\max}(\boldsymbol{X}_j) \leq R \text{ almost surely.}$$

Compute the norm of total variance,

$$\sigma^2 := \left\|\sum_j \mathbb{E}[\boldsymbol{X}_j^2]\right\|_2.$$

Then the following inequality holds for all $\tau \geq 0$*:*

$$\mathbb{P}\left\{\lambda_{\max}\Big(\sum_j \boldsymbol{X}_j\Big) \geq \tau\right\} \leq \begin{cases} n\exp(-3\tau^2/(8\sigma^2)), \ \tau \leq \sigma^2/R, \\ n\exp(-3\tau/(8R)), \ \tau > \sigma^2/R. \end{cases}$$

Using Theorem 2, we can prove the following lemma.

Lemma 2 *Let* $\varepsilon > 0$*, and* $L \in (0,1)$*,* $k = \Omega(\log(\frac{L}{n^{1.5}})/\log(\lambda_2))$*,* v^2 *be the variance defined in Theorem 1 and assume* $v^2 > 0$*, then we have*

$$\mathbb{P}\left\{\|\boldsymbol{C} - \hat{\boldsymbol{C}}\|_2 \geq \varepsilon\|\boldsymbol{C}\|_2\right\} \leq \begin{cases} 2n\exp\left(-\frac{3N\lambda_1^2\varepsilon^2}{8v^2}\right), \ \text{if } \varepsilon \leq v^2/(1+L)^2, \\ 2n\exp\left(-\frac{3N\lambda_1\varepsilon}{8(1+L)^2}\right), \ \text{if } \varepsilon > v^2/(1+L)^2, \end{cases}$$

where λ_1, λ_2 *are 2 dominant eigenvalues of* $\boldsymbol{W}$ *and* $\|\boldsymbol{C}\|_2 = \lambda_1 = 1$*.*

Proof $\forall j \in [N]$, let $\boldsymbol{Y}_j := n(\boldsymbol{W}^k\boldsymbol{e}_{i_j})(\boldsymbol{W}^k\boldsymbol{e}_{i_j})^T$, then we have

$$\hat{\boldsymbol{C}} = \frac{n}{N}\boldsymbol{B}\boldsymbol{B}^T = \frac{1}{N}\sum_{j=1}^{N}\boldsymbol{Y}_j,$$

and for d-$\boldsymbol{W} = \boldsymbol{W}^T$, therefore

$$\begin{aligned} \mathbb{E}[\boldsymbol{Y}_j] &= n\sum_{l=1}^{n}\mathbb{P}[i_j = l]\boldsymbol{W}^k\boldsymbol{e}_l\boldsymbol{e}_l^T(\boldsymbol{W}^k)^T \\ &= \boldsymbol{W}^k\left(\sum_{l=1}^{n}\boldsymbol{e}_l\boldsymbol{e}_l^T\right)\boldsymbol{W}^{kT} \\ &= \boldsymbol{W}^k\boldsymbol{I}\boldsymbol{W}^{kT} \\ &= \boldsymbol{W}^k\boldsymbol{W}^{kT} = \boldsymbol{C} \end{aligned}$$

Observe that

$$\begin{aligned} \mathbb{P}\left\{\|\boldsymbol{C} - \hat{\boldsymbol{C}}\| \geq t\right\} &= \mathbb{P}\left\{\lambda_{\max}(\boldsymbol{C} - \hat{\boldsymbol{C}}) \geq t \text{ or } \lambda_{\max}(\hat{\boldsymbol{C}} - \boldsymbol{C}) \geq t\right\} \\ &\leq \mathbb{P}\left\{\lambda_{\max}(\boldsymbol{C} - \hat{\boldsymbol{C}}) \geq t\right\} + \mathbb{P}\left\{\lambda_{\max}(\hat{\boldsymbol{C}} - \boldsymbol{C}) \geq t\right\} \\ &= \mathbb{P}\left\{\lambda_{\max}\Big(\sum_{i=1}^{N}(\boldsymbol{C} - \boldsymbol{Y}_j)\Big) \geq Nt\right\} \\ &\quad + \mathbb{P}\left\{\lambda_{\max}\Big(\sum_{i=1}^{N}(\boldsymbol{Y}_j - \boldsymbol{C})\Big) \geq Nt\right\}. \end{aligned} \tag{3}$$

Note that both $\mathbb{E}[\boldsymbol{C} - \boldsymbol{Y}_j] = \mathbb{E}[\boldsymbol{Y}_j - \boldsymbol{C}] = \boldsymbol{0}$. Since $\boldsymbol{C}$ and $\boldsymbol{Y}_j$ are both positive semi-definite, we have

$$\lambda_{\max}(\boldsymbol{C} - \boldsymbol{Y}_j) = \max_{\boldsymbol{v}:\|\boldsymbol{v}\|=1} \boldsymbol{v}^T(\boldsymbol{C} - \boldsymbol{Y}_j)\boldsymbol{v} \leq \lambda_{\max}(\boldsymbol{C}) = 1,$$

and by our choice of L, k and Lemma 1, we have

$$\begin{aligned}\lambda_{\max}(\boldsymbol{Y}_j - \boldsymbol{C}) &= \max_{\boldsymbol{v}:\|\boldsymbol{v}\|=1} \boldsymbol{v}^T(\boldsymbol{Y}_j - \boldsymbol{C})\boldsymbol{v} \leq \max_{\|\boldsymbol{v}\|=1} n(\boldsymbol{v}^T \boldsymbol{W}^k \boldsymbol{e}_{i_j})^2 \\ &\leq \max_{\boldsymbol{v}:\|\boldsymbol{v}\|=1} n\|\boldsymbol{v}\|^2 \|\boldsymbol{W}^k \boldsymbol{e}_{i_j}\|^2 \leq (1+L)^2.\end{aligned}$$

Thus the upper bound R in Theorem 2 is $(1+L)^2$. The variance parameter σ^2 is

$$\begin{aligned}\sigma^2 &= \left\| \sum_{j=1}^{N} \sum_{l=1}^{n} \mathbb{P}[i_j = l] \left(n(\boldsymbol{W}^k \boldsymbol{e}_l)(\boldsymbol{W}^k \boldsymbol{e}_l)^T - \boldsymbol{C} \right)^2 \right\| \\ &= N \left\| \frac{1}{n} \sum_{i=1}^{n} \left(n(\boldsymbol{W}^k \boldsymbol{e}_i)(\boldsymbol{W}^k \boldsymbol{e}_i)^T - \boldsymbol{C} \right)^2 \right\| = N\nu^2.\end{aligned}$$

Now assume $\varepsilon \leq \frac{\nu^2}{\lambda_1(1+L)^2}$ and let $t = \lambda_1 \varepsilon = \varepsilon \|\boldsymbol{C}\|_2$, we have $Nt \leq \frac{N\nu^2}{(1+L)^2} = \frac{\sigma^2}{R}$. Applying upper branch of Theorem 2 to Eq. 3, we get the desired upper branch. Similarly, when $\varepsilon > \frac{\nu^2}{\lambda_1(1+L)^2}$ and let $t = \lambda_1 \varepsilon = \varepsilon \|\boldsymbol{C}\|_2$, we have $Nt > \frac{\sigma^2}{R}$. By applying lower branch of Theorem 2 to Eq. 3, we get the desired lower branch. □

Now we are ready to prove Theorem 1.

Proof of Theorem 1 Assume $\varepsilon \leq \frac{\nu^2}{(1+L)^2}$, and let β be a parameter that controls the high probability result, then

$$N \geq \frac{8}{3}(1+\beta)\left(\frac{\nu^2}{\varepsilon^2} \log(2n)\right)$$

implies

$$\mathbb{P}\left\{ \|\boldsymbol{C} - \hat{\boldsymbol{C}}\| \geq \varepsilon \|\boldsymbol{C}\| \right\} \leq 2n \exp\left(-\frac{3N\varepsilon^2}{8\nu^2} \right) \leq (2n)^{-\beta}.$$

Otherwise if $\varepsilon > \frac{\nu^2}{(1+L)^2}$, then

$$N \geq \frac{8}{3}(1+\beta)\left(\frac{(1+L)^2}{\varepsilon} \log(2n)\right)$$

also implies

$$\mathbb{P}\left\{\|\boldsymbol{C}-\hat{\boldsymbol{C}}\| \geq \varepsilon\|\boldsymbol{C}\|\right\} \leq 2n \exp\left(-\frac{3N\varepsilon}{8(1+L)^2}\right) \leq (2n)^{-\beta}.$$

□

Next we extend our error bound on $\|\boldsymbol{C}-\hat{\boldsymbol{C}}\|$ to an error bound on $\|\boldsymbol{M}-\boldsymbol{C}\|$. To achieve this, we show that $\boldsymbol{M}$ is actually a good approximation of $\hat{\boldsymbol{C}}$ when α is close to 1 and k is reasonably large. Formally, we define $\boldsymbol{R}$ as

$$\left[\frac{\log.(\boldsymbol{x}(i_1,\alpha))}{\|\log.(\boldsymbol{x}(i_1,\alpha))\|} \ \frac{\log.(\boldsymbol{x}(i_2,\alpha))}{\|\log.(\boldsymbol{x}(i_2,\alpha))\|} \ \cdots \ \frac{\log.(\boldsymbol{x}(i_N,\alpha))}{\|\log.(\boldsymbol{x}(i_N,\alpha))\|}\right],$$

and let $\boldsymbol{M}$ be $\frac{1}{N}\boldsymbol{R}\boldsymbol{R}^T$.

As $\lim_{\alpha\to 1^-}\log.(\boldsymbol{x}(u,\alpha)) = \log.(\boldsymbol{\pi}) = \log.(\frac{\boldsymbol{e}}{n})$ for d-regular graphs, we assume that

$$\left\|\frac{\log.(\boldsymbol{x}(u,\alpha))}{\|\log.(\boldsymbol{x}(u,\alpha))\|} - \frac{\boldsymbol{e}}{\sqrt{n}}\right\| \leq \gamma_\alpha, u \in V$$

where

$$\lim_{\alpha\to 1^-}\gamma_\alpha = 0.$$

With these quantities defined, we have the following theorem.

Theorem 3 *Let $\tau \in (0,1)$, $L \leq \frac{\tau}{6}$, we choose α close to 1 such that $\gamma_\alpha \leq \frac{\tau}{6}$ and pick k according to Theorem 1, then we have*

$$\|\boldsymbol{M}-\hat{\boldsymbol{C}}\| \leq \tau.$$

To prove it, we first prove the following lemma.

Lemma 3 *For L, k defined as in Theorem 1, let $\boldsymbol{y}_j = \sqrt{n}\boldsymbol{W}^k\boldsymbol{e}_{i_j}$ and $\boldsymbol{z}_j = \frac{\log.(\boldsymbol{x}(i_j,\alpha))}{\|\log.(\boldsymbol{x}(i_j,\alpha))\|}$, we have*

$$\|\boldsymbol{y}_j\boldsymbol{y}_j^T - \boldsymbol{z}_j\boldsymbol{z}_j^T\| \leq 2(L+\gamma_\alpha) + L^2 + \gamma_\alpha^2.$$

Proof Let $\boldsymbol{\epsilon}_y = \boldsymbol{y}_j - \frac{\boldsymbol{e}}{\sqrt{n}}$ and $\boldsymbol{\epsilon}_z = \boldsymbol{z}_j - \frac{\boldsymbol{e}}{\sqrt{n}}$, by Lemma 1 and our assumption above, we have

$$\|\boldsymbol{\epsilon}_y\| \leq L, \|\boldsymbol{\epsilon}_z\| \leq \gamma_\alpha.$$

By definition of matrix 2-norm, we know that

$$\|\boldsymbol{y}_j\boldsymbol{y}_j^T - \boldsymbol{z}_j\boldsymbol{z}_j^T\|$$

$$
\begin{aligned}
&= \max_{\boldsymbol{v}:\|\boldsymbol{v}\|=1} \|\boldsymbol{v}^T \left(\boldsymbol{y}_j \boldsymbol{y}_j^T - \boldsymbol{z}_j \boldsymbol{z}_j^T \right) \boldsymbol{v}\| \\
&= \max_{\boldsymbol{v}:\|\boldsymbol{v}\|=1} \|\boldsymbol{v}^T \left((\frac{\boldsymbol{e}}{\sqrt{n}} + \boldsymbol{\epsilon}_y)(\frac{\boldsymbol{e}}{\sqrt{n}} + \boldsymbol{\epsilon}_y)^T - (\frac{\boldsymbol{e}}{\sqrt{n}} + \boldsymbol{\epsilon}_z)(\frac{\boldsymbol{e}}{\sqrt{n}} + \boldsymbol{\epsilon}_z)^T \right) \boldsymbol{v}\| \\
&\leq \max_{\boldsymbol{v}:\|\boldsymbol{v}\|=1} (\|\boldsymbol{v}\|\|\boldsymbol{\epsilon}_y\|)^2 + 2\|\boldsymbol{v}\|(\|\boldsymbol{v}\|\|\boldsymbol{\epsilon}_y\|) + (\|\boldsymbol{v}\|\|\boldsymbol{\epsilon}_z\|)^2 + 2\|\boldsymbol{v}\|(\|\boldsymbol{v}\|\|\boldsymbol{\epsilon}_z\|) \\
&\leq 2(L + \gamma_\alpha) + L^2 + \gamma_\alpha^2.
\end{aligned}
$$

□

Now we apply it to prove Theorem 3.

Proof of Theorem 3 For $\forall j \in [N]$, define $\boldsymbol{y}_j, \boldsymbol{z}_j$ according to Lemma 3, and let $\boldsymbol{Y}_j = n\boldsymbol{y}_j^T \boldsymbol{y}_j$, $\boldsymbol{Z}_j = \boldsymbol{z}_j \boldsymbol{z}_j^T$, then we have

$$
\boldsymbol{M} - \hat{\boldsymbol{C}} = \frac{1}{N} \left(\sum_{j=1}^{N} \boldsymbol{Z}_j - \sum_{j=1}^{N} \boldsymbol{Y}_j \right).
$$

Thus by Lemma 3, we have

$$
\begin{aligned}
\|\boldsymbol{M} - \hat{\boldsymbol{C}}\| &\leq \frac{1}{N} \sum_{i=1}^{N} \|\boldsymbol{y}_j \boldsymbol{y}_j^T - \boldsymbol{z}_j \boldsymbol{z}_j^T\| \\
&\leq 2(L + \gamma_\alpha) + L^2 + \gamma_\alpha^2 \\
&\leq \tau.
\end{aligned}
$$

□

Theorems 1 and 3 empower us to show the following result with regard to the relation between dominant singular vectors of $\boldsymbol{R}$ and eigenvectors of $\boldsymbol{W}$.

Theorem 4 *Let $\varepsilon > 0, \tau \in (0, 1)$ and choose L, k, N, α according to Theorems 1 and 3, then we have*

$$
\|\boldsymbol{M} - \boldsymbol{C}\| \leq \tau + \varepsilon,
$$

and denote by $\boldsymbol{U}_l \in \mathbb{R}^{n \times l}$ the l dominant eigenvectors of $\boldsymbol{C}$, $\hat{\boldsymbol{U}}_l \in \mathbb{R}^{n \times l}$ the l dominant eigenvectors of $\boldsymbol{M}$, we have

$$
\lambda_l(\boldsymbol{C}) \|\sin \Theta(\boldsymbol{U}_l, \hat{\boldsymbol{U}}_l)\| \leq \|(\boldsymbol{I} - \boldsymbol{U}_l \boldsymbol{U}_l^T)\boldsymbol{C}\| + 2(\tau + \varepsilon).
$$

Before proving Theorem 4, we introduce two useful theorems from Drineas and Ipsen (2019).

Theorem 5 (Corollary 2 of Drineas and Ipsen (2019)) *Let $\boldsymbol{U}_k \in \mathbb{R}^{n\times k}$ be k dominant left singular vectors of $\boldsymbol{A}$; and let $\hat{\boldsymbol{U}}_k \in \mathbb{R}^{m\times k}$ be k dominant left singular vectors of $\boldsymbol{A}+\boldsymbol{E}$. Then*

$$\|(\boldsymbol{I}-\boldsymbol{U}_k\boldsymbol{U}_k^T)\boldsymbol{A}\|_2 \le \|(\boldsymbol{I}-\hat{\boldsymbol{U}}_k\hat{\boldsymbol{U}}_k)\boldsymbol{A}\|_2 \le \|(\boldsymbol{I}-\boldsymbol{U}_k\boldsymbol{U}_k^T)\boldsymbol{A}\|_2 + 2\|\boldsymbol{E}\|_2.$$

Theorem 6 (Theorem 6 of Drineas and Ipsen (2019)) *Let $\boldsymbol{P}_k \equiv \boldsymbol{A}_k\boldsymbol{A}_k^{\dagger}$ be the orthogonal projector ($\boldsymbol{P}^2=\boldsymbol{P}=\boldsymbol{P}^T$) onto the dominant k-dimensional subspace of $\boldsymbol{A}$; and let $\boldsymbol{P}\in\mathbf{R}^{m\times m}$ with $k \le \operatorname{rank}(\boldsymbol{P}) < m-k$. Then*

$$\sigma_k(\boldsymbol{A})|||\sin\Theta(\boldsymbol{P},\boldsymbol{P}_k)|||_p \le |||(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{A}|||_p \le \|\boldsymbol{A}\|_2|||\sin\Theta(\boldsymbol{P},\boldsymbol{P}_k)|||_p + |||\boldsymbol{A}-\boldsymbol{A}_k|||_p,$$

where $|||\cdot|||$ denotes Schatten p-norms.

Now we can prove Theorem 4 using these two theorems.

Proof of Theorem 4 By our choice of L, k, N, α, we have

$$\|\boldsymbol{M}-\boldsymbol{C}\| \le \|\boldsymbol{M}-\hat{\boldsymbol{C}}\| + \|\boldsymbol{C}-\hat{\boldsymbol{C}}\| \le \tau+\varepsilon.$$

Further by Theorem 5, we know

$$\begin{aligned}\|(\boldsymbol{I}-\hat{\boldsymbol{U}}_l\hat{\boldsymbol{U}}_l^T)\boldsymbol{C}\| &\le \|(\boldsymbol{I}-\boldsymbol{U}_l\boldsymbol{U}_l^T)\boldsymbol{C}\| + 2\|\boldsymbol{M}-\boldsymbol{C}\| \\ &\le \|(\boldsymbol{I}-\boldsymbol{U}_l\boldsymbol{U}_l^T)\boldsymbol{C}\| + 2(\tau+\varepsilon).\end{aligned}$$

Thus using Theorem 6, we get

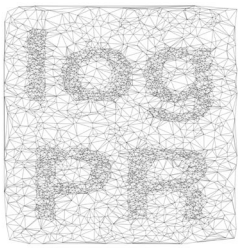
$$\begin{aligned}\lambda_l(\boldsymbol{C})\|\sin\Theta(\boldsymbol{U}_l,\hat{\boldsymbol{U}}_l)\| &\le \|(\boldsymbol{I}-\hat{\boldsymbol{U}}_l\hat{\boldsymbol{U}}_l^T)\boldsymbol{C}\| \\ &\le \|(\boldsymbol{I}-\boldsymbol{U}_l\boldsymbol{U}_l^T)\boldsymbol{C}\| + 2(\tau+\varepsilon).\end{aligned}$$

□

Because $\boldsymbol{C}$ is the Gram matrix of $\boldsymbol{W}$ and $\boldsymbol{M}$ is the scaled Gram matrix of $\boldsymbol{R}$, we know that the eigenvectors of $\boldsymbol{M}$ correspond to the left singular vectors of $\boldsymbol{R}$ and the eigenvectors of $\boldsymbol{C}$ correspond to the eigenvectors of $\boldsymbol{W}$. So the theorem above also characterizes the angles between the subspaces spanned by the dominant left singular vectors of $\boldsymbol{R}$ and the dominant eigenvectors of $\boldsymbol{W}$.

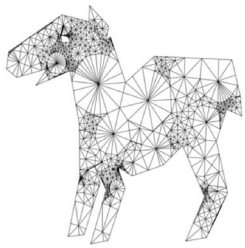
6 Empirical comparison results

We study the log-PageRank embedding on synthetic and real world graphs. We focus on those where the spectral embedding gives a good picture of the graph, as spectral embeddings may fail to give useful pictures for many real-world networks (Lang 2005) and Fig. 2. We use the following four classes of graphs.



(a) Graph for the word "log PR" with 5000 nodes, 14962 edges

(b) The Minnesota road network with 2640 nodes, 3302 edges



(c) The Tapir (Bern et al, 1994) graph with 1024 nodes, 2846 edges

Fig. 6 Our geometric graphs

Nearest Neighbor Graphs

For a graph named n-k nearest neighbor, there are n points randomly distributed in the unit square and connected to k nearest neighbors.

Chain Graphs

These are simply the chain graphs we have from the analysis in Sect. 4.

Graphs with Strong Geometry

These are the graphs from Fig. 6.

Stochastic Block Models

A graph named sbm(n, k, p, q) has k groups of n vertices with in-block edge probability, p, and out-block edge probability, q. These show the worst approximation results for one metric and largest differences. This is due to a nearly multiple dimensional eigenspace.

6.1 Implementation

We implement Algorithm 1 for log-PageRank embedding in Julia. The built in sparse LU solver was used to factorize the PageRank matrix $\boldsymbol{I} - \alpha \boldsymbol{P}$ to solve linear systems for large values of α and the built in dense SVD solver for the embeddings. There are many alternatives one could use here, but our focus was on understanding the log-PageRank embeddings rather than optimizing the speed at which those can be computed. We did not work to optimize computational runtime as it was not limiting our experiments. Ideas related to streaming SVD computation (Constantine and Gleich 2011; Tsitsulin et al. 2021) may be used if SVD computation is problematic. Any of the extensive literature on fast PageRank computation may also be used if PageRank becomes a bottleneck.

6.2 Qualitative approximation

Given a spectral embedding and a log-PageRank embedding, we first simply look at the differences in the pictures. Our results say these should look similar, not that they should be exactly the same as the graphs we study are not in the class where we expect

sharp approximations. Let the second singular vector of the log-PageRank embedding be $\boldsymbol{u}_2$, and the second eigenvector of the Laplacian be $\boldsymbol{z}_2$. Likewise for the 3rd vectors. So the spectral embedding is $\boldsymbol{z}_2$, $\boldsymbol{z}_3$ and the log-PageRank embedding is $\boldsymbol{u}_2$, $\boldsymbol{u}_3$.

The first way we evaluate the embeddings is by looking at the joint plot of $\boldsymbol{u}_2$ vs. $\boldsymbol{z}_2$ and $\mathbf{u}_3$ vs. $\boldsymbol{z}_3$. If the embeddings are close, these should look like a straight line, or at least a very highly correlated relationships.

Figures 7 and 8 show the embeddings as α varies both with and without the nonlinear *log* operation for the graphs in Fig. 6b, c. The result on the "log PR" graph from Fig. 6a is in the introduction.

On nearest neighbor graphs, such as Fig. 9, these embeddings show a clear rotational ambiguity that might arise with other evaluations of this strategy. (This occured with the other graphs too.) Put plainly, the eigenvectors are almost in 2d invariant subspace. Consequently, when we randomize the method, we can only capture this near 2d subspace up to rotation. However, this will not show high error with respect to the approximation error measure as the results are all near eigenvectors.

We show one of the examples of the stochastic block model in Fig. 10. Although this has bad approximation with respect to the spectral embedding, the result for the log-PageRank embedding for $\alpha = 0.99$ is arguably better than the spectral embedding.

6.3 Quantitative error and approximation

We quantitatively measure the error in three ways.

Rayleigh quotient
The first way is by evaluating the relative difference between the Rayleigh quotient with respect to the first dimension used for embedding, i.e., the second singular vectors of the log-PageRank matrix $\boldsymbol{u}_2$ and the 2nd smallest eigenvector of the Laplacian $\boldsymbol{z}_2$. Let

$$s_2 = \frac{\boldsymbol{z}_2^T \boldsymbol{\mathcal{L}} \boldsymbol{z}_2}{\boldsymbol{z}_2^T \boldsymbol{z}_2}, \qquad p_2 = \frac{\boldsymbol{u}_2^T \boldsymbol{\mathcal{L}} \boldsymbol{u}_2}{\boldsymbol{u}_2^T \boldsymbol{u}_2}$$

then we have the following measure:

$$R_2 = \left| \frac{s_2 - p_2}{s_2} \right|. \tag{4}$$

Multiple Rayleigh quotients
We also evaluate with the information from the 3rd singular vector according to our error measure. Let

$$s_{23} = \frac{\boldsymbol{z}_2^T \boldsymbol{L} \boldsymbol{z}_2}{\boldsymbol{z}_2^T \boldsymbol{z}_2} + \frac{\boldsymbol{z}_3^T \boldsymbol{L} \boldsymbol{z}_3}{\boldsymbol{z}_3^T \boldsymbol{z}_3}, \qquad p_{23} = \frac{\boldsymbol{u}_2^T \boldsymbol{L} \boldsymbol{u}_2}{\boldsymbol{u}_2^T \boldsymbol{u}_2} + \frac{\boldsymbol{u}_3^T \boldsymbol{L} \boldsymbol{u}_3}{\boldsymbol{u}_3^T \boldsymbol{u}_3}.$$

Table 2 Error between PageRank embedding and spectral embedding for different graphs at a low teleportation probability, $\alpha = 0.99$ and at a higher one $\alpha = 0.9999$ both without log (raw) and with log using only the second vector as in (4) (lower is better)

Graph	$\alpha = 0.99$		$\alpha = 0.9999$	
	Raw (%)	Log (%)	Raw (%)	Log (%)
30-6 nearest neighbour	3.18	2.02	2.45	4.05
3000-6 nearest neighbour	47.6	0.37	5.28	2.57
10000-6 nearest neighbour	169.75	2.13	14.96	1.55
30 chain	16.65	0.51	25.88	4.57
3000 chain	5556.36	2.17	57.75	2.36
Minnesota $n = 2640$	16.34	1.93	11.35	0.84
Tapir $n = 1024$	10.17	1.13	15.3	0.73
LogPR $n = 5000$	19.95	0.15	4.76	0.34
sbm(50,60,0.001,0.005)	53.99	19.87	53.57	68.55
sbm(1000,3,0.001,0.005)	56.07	20.11	55.47	92.03
sbm(50,60,0.25,0.005)	93.02	29.69	92.59	100.48
sbm(1000,3,0.25,0.001)	453.32	5.51	140.23	105.33

then we look at the difference

$$R_{23} = \left| \frac{s_{23} - p_{23}}{s_{23}} \right|. \tag{5}$$

Subspace error
Finally, we evaluate the difference between subspaces. A related measure was also used in Tsitsulin et al. (2021). Let $\boldsymbol{U} = \left[\boldsymbol{u}_2\ \boldsymbol{u}_3\right]$ be the subspace from the log-PageRank embedding and $\boldsymbol{Z} = \left[z_2\ z_3\right]$ be the subspace the spectral embedding. Then the distance (or gap) between $\boldsymbol{U}$ and $\boldsymbol{Z}$ is

$$C = \text{dist}(\text{range}(\boldsymbol{U}), \text{range}(\boldsymbol{Z})) = \|\boldsymbol{U}\boldsymbol{U}^T - \boldsymbol{Z}\boldsymbol{Z}^T\|_2. \tag{6}$$

See Stewart (1973) for this error measure.
Results
The above three error measures are evaluated in Tables 2 (for R_2), 3 (for R_{23}), and 4 (for C). In virtually all of the experiments, the log-PageRank embedding has a strikingly lower error measure than the same embedding without the log. When this is not the case, such as on the SBM graphs with a planted partition, the problems are known to have nearly duplicate eigenvectors.

6.4 Embedding error variance

We study the dependence of this error on the number of randomly sampled nodes and location of sampled nodes in the graph. We record this for log-PageRank at $\alpha = 0.99$

Table 3 Error between PageRank embedding and spectral embedding for different graphs at a low teleportation probability, $\alpha = 0.99$ and at a higher one $\alpha = 0.9999$ both without log (raw) and with log using both the second and third vector as in (5) (lower is better)

Graph	$\alpha = 0.99$		$\alpha = 0.9999$	
	Raw (%)	Log (%)	Raw (%)	Log (%)
30-6 nearest neighbour	15.7	0.07	13.46	11.72
3000-6 nearest neighbour	50.67	1.01	9.8	4.86
10000-6 nearest neighbour	148.48	0.03	10.65	0.41
30 chain	6.31	11.46	12.54	2.74
3000 chain	5648.6	1.22	53.53	7.24
Minnesota $n = 2640$	7.38	0.26	6.9	0.39
Tapir $n = 1024$	6.9	1.37	10.41	0.82
LogPR $n = 5000$	22.07	0.19	4.37	0.12
sbm(50,60,0.001,0.005)	56.69	19.64	56.29	67.78
sbm(1000,3,0.001,0.005)	62.45	19.04	61.85	90
sbm(50,60,0.25,0.005)	93.42	31.81	93.01	99.05
sbm(1000,3,0.25,0.001)	594.21	7.31	189.95	140.44

Table 4 Error between PageRank embedding and spectral embedding for different graphs at a low teleportation probability, $\alpha = 0.99$ and at a higher one $\alpha = 0.9999$ both without log (raw) and with log using column covariance error definition as in (6) (lower is better)

Graph	$\alpha = 0.99$		$\alpha = 0.9999$	
	Raw (%)	Log (%)	Raw (%)	Log (%)
30-6 nearest neighbour	49.8	20.7	43.9	23.1
3000-6 nearest neighbour	145.66	17.55	38.68	12.5
10000-6 nearest neighbour	187.16	20.51	43.56	18.11
30 chain	57.71	30.2	43.03	26.9
3000 chain	196.41	26.4	161.4	21.27
Minnesota $n = 2640$	179.9	75.5	74.81	58.23
Tapir $n = 1024$	87.2	30.12	74.12	9.41
LogPR $n = 5000$	105.8	23.3	45.84	9.58
sbm(50,60,0.001,0.005)	199.6	198.8	199.64	199.68
sbm(1000,3,0.001,0.005)	199.5	198.4	199.4826	199.4829
sbm(50,60,0.25,0.005)	198.43	188.33	198.37	199.9
sbm(1000,3,0.25,0.001)	21.6	2.3	14.4	8.25

in Fig. 11. This shows the distribution of errors as a density estimate, along with the max/min values (small) and the median value (big).

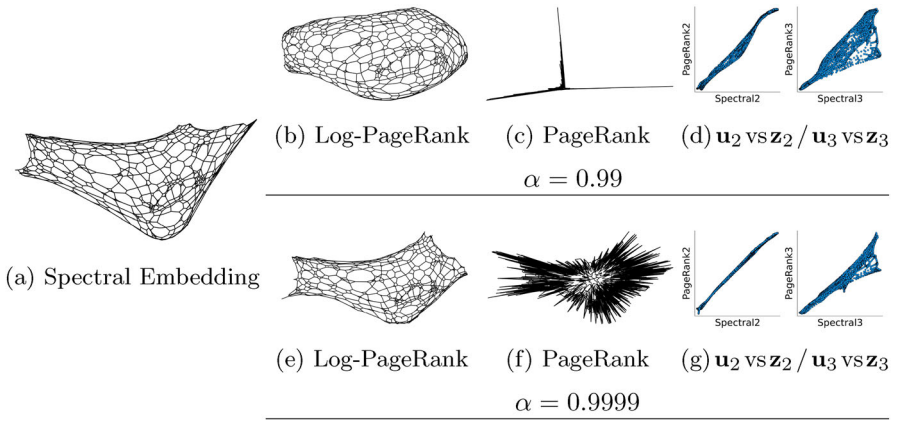


Fig. 7 Comparison of embeddings for the Minnesota network

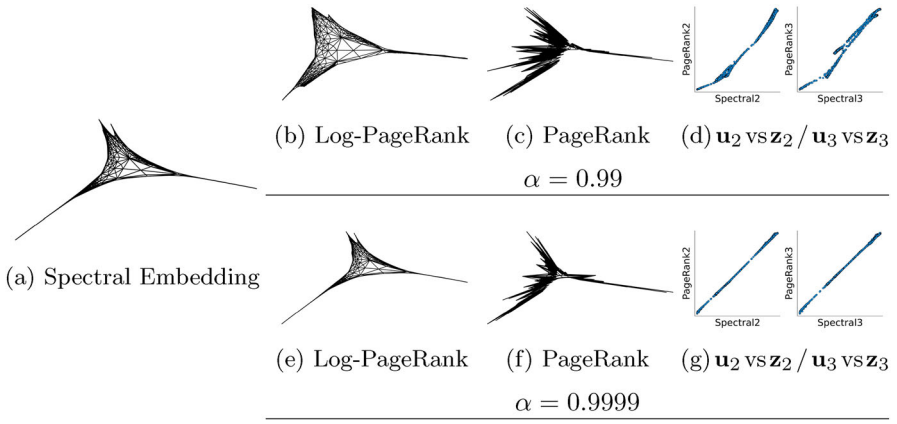


Fig. 8 Comparison of embeddings for the Tapir graph

As expected, there are largely minimal effects. This occurs because the sensitivity of PageRank to the seed vector, $\mathbf{v}$, is a function of α (Langville and Meyer 2006).

$$\frac{d\boldsymbol{x}}{d\boldsymbol{v}} = (1-\alpha)(\boldsymbol{I} - \alpha \boldsymbol{P})^{-1}$$

which satisfies $\|\frac{d\boldsymbol{x}}{d\boldsymbol{v}}\|_1 = 1$. Further, for $\alpha \to 1$, dependence of the PageRank values on $\mathbf{v}$ reduces. Our experiments confirm the same as the minimum, maximum and variance of error over 50 trials show negligible change.

7 Hypergraph embeddings

One driving reason for our study of the log-PageRank embedding is to support similar embedding strategies for different types of data, such as those studied in Fountoulakis

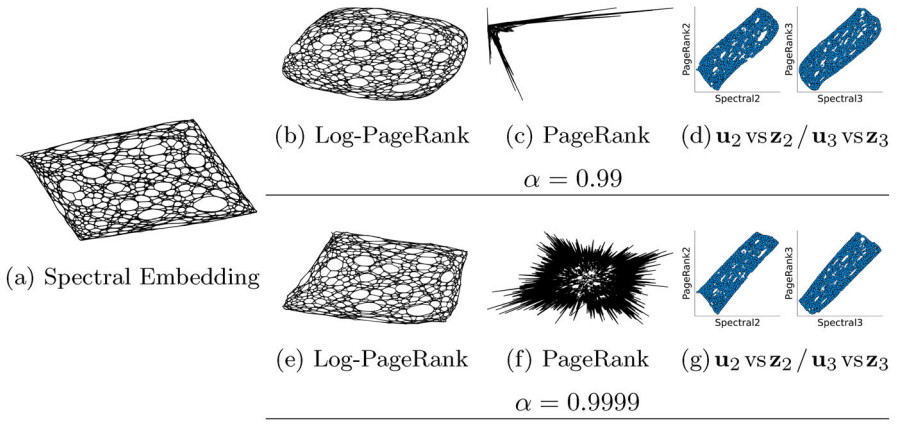


Fig. 9 Embedding for 10000 node graph with 6 nearest neighbours. Note that log-PageRank and spectral embeddings are fairly similar after a rotation. This is expected because the corresponding eigenvalues are close in magnitude

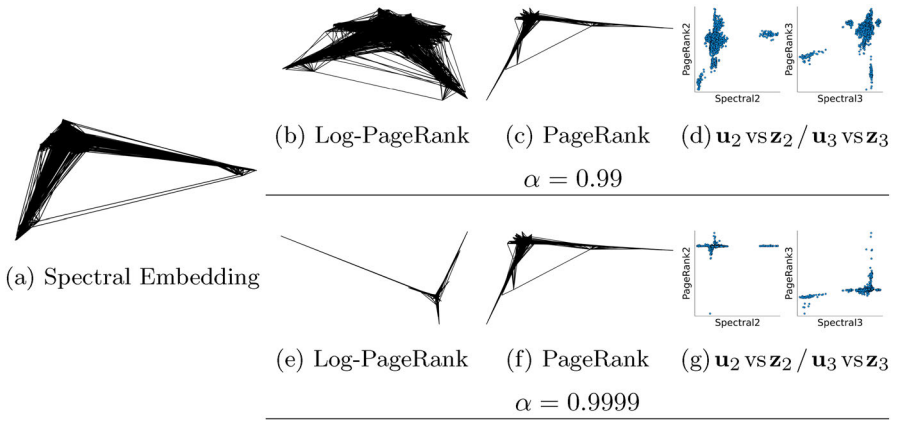


Fig. 10 Comparison of the embedding techniques on the planted partition model with good conductance cuts. The model here is (60 blocks of 50 nodes each with in block edge probability of 0.25 and out-block edge probability of 0.001). This is one case where the technique does not seem to work. This is likely due to the highly degenerate eigenspace created by the stochastic block models

et al. (2020). In this section, we use the log-PageRank embedding technique on five hypergraphs: Yelp (https://www.yelp.com/dataset), Walmart Trips (Amburg et al. 2020), a contact tracing network (Benson et al. 2018; Stehlé et al. 2011), posts on Math Overflow (Veldt et al. 2020a), and a Drug Abuse network (DAWN) (Amburg et al. 2020). The only modification to Algorithm 1 is that we replace seeded PageRank with the Local Quadratic PageRank, a method proposed in Liu et al. (2021). Specifically we use the LQHD method with a 2-norm penalty with $\rho = 0.5$ for all experiments. For the Yelp and Walmart trips network, we set $\kappa = 0.000025$ and $\gamma = 1.0$ while for the Math Overflow network, with the same sparsity factor $\kappa = 0.000025$, we set $\gamma = 0.001$. For Contact Primary School and DAWN, we set $\kappa = 0.0025$ and $\gamma = 0.001$. These

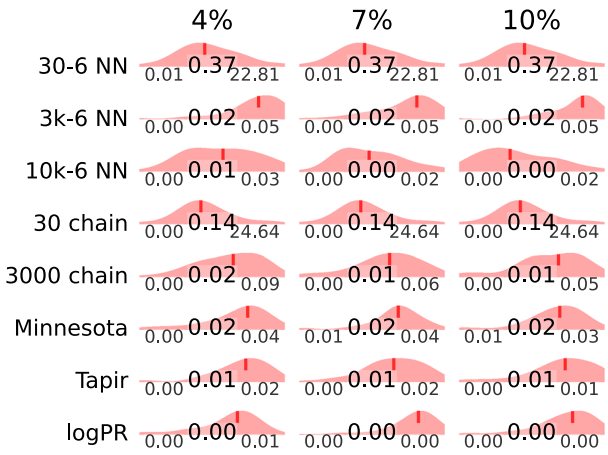


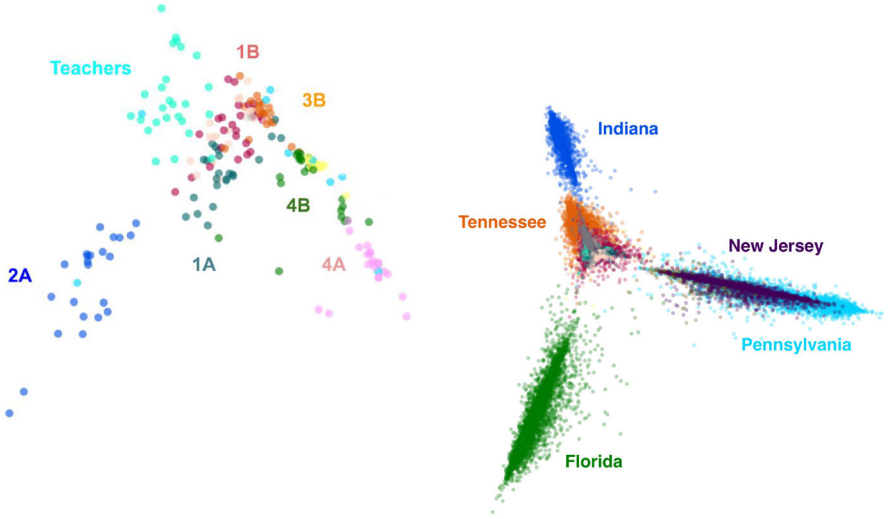
Fig. 11 Error variation with column for log of PageRank with $\alpha = 0.99$. The percentage indicated in the column headings are the fraction of the nodes as seeds. Each entry is the variance, the maximum and the minimum for 50 trials

choices were made arbitrarily, there are small differences that result when changing them.

Figure 12a shows our 2d embedding on Contact Primary School dataset where each node represents a student or a teacher, each hyperedge represents a group of people who are spatially close at a given time. Each node of the graph is colored as a teacher or as classroom for the student. Note that each classroom forms a cohesive group in the plot. Moreover, we observe that the students from the same grade, e.g. students colored red (1B) and dark green (1A), share some spatial proximity, which is due to the fact that their classrooms are close. One notable observation is that teachers do form a group in the embedding, who are mostly separated from the students.

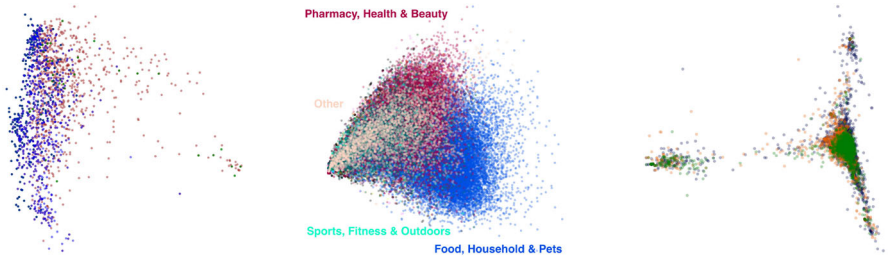
Figure 12b shows our embedding on Yelp Review data. Following Veldt et al. (2020b), we build one hypergraph with each restaurant being a node and each user being a hyperedge. We show the state associated with each location as the color. We can clearly see that our embedding captures the geographic information of the underlying hypergraph. For example, the nodes labeled dark blue are those restaurants from state Indiana, which are close to the orange nodes from state Tennessee. Further, the green nodes from state Florida are quite well-separated from nodes with other colors, which is due to the fact that none of other 13 states (Pennsylvania, Tennessee, Missouri, Indiana, Alabama, Nevada, Illinois, Arizona, Louisiana, New Jersey, California, Delaware, Idaho) we plot is close to Florida.

In addition, we show log-PageRank embeddings of three other hypergraphs in Fig. 13a–c. We are unable to identify obvious relationships between these embeddings and the existing groups, which means the embeddings likely show a different type of structure. Notably, the obvious product category partitions of the Walmart data are not reflected in the layout structure. The promising results on all the datasets above show that our simple algorithm is capable of generating good embeddings even on higher order graphs.



(a) Contact Primary School (b) Yelp Restaurants

Fig. 12 Log-PageRank embedding of hypergraphs. The Contact Primary School dataset has 242 nodes and 12704 hyperedges. Nodes are colored by classroom and teachers, which form cohesive groups due to the contact structure. The Yelp Restaurant dataset has 52260 nodes and 597261 hyperedges. Nodes are colored by one of 14 states used for analysis, which show clear geographic relationships



(a) DAWN (b) Walmart Trips (c) Math Overflow

Fig. 13 Log-PageRank embeddings of the (**a**) DAWN dataset (Amburg et al. 2020) with 2109 nodes (drugs) and 87104 hyperedges where each hyperedge is an individual and the consisting nodes are the drugs consumed by them; the plot shows 3 out 10 labels (**b**) the Walmart dataset (Amburg et al. 2020) with 88860 nodes and 69906 hyperedges where each node is a product and each hyperedge consists of products purchased in one trip to Walmart and (**c**) the Math Overflow (Veldt et al. 2020a) dataset with 73851 nodes, 5446 hyperedges, where each hyperedge shows the multiple labels associated with a question; Although there is structure evident in the plots, it does not strongly correlate with the known labels on the data (some of the plotting makes the structure look more present than it is)

8 Related research to log PageRank

The idea of using the log of a PageRank vector originated in Google's initial use of these for their PageRank scores and their use in spam ranking as discussed in the introduction. Since then, the elementwise log of PageRank values emerged in other scenarios as well. The most recent is its use in graph embeddings. A key task in graph embeddings is to sample a set of related nodes and then fit a lower dimensional embedding vector for each node where related nodes are close and nodes that were not sampled as related are far. These ideas build on word embeddings techniques such as SkipGram models (Grover and Leskovec 2016; Tang et al. 2015a, b; Perozzi et al. 2014). One technique to sample related entries is to use a seeded PageRank random walk (Zhou et al. 2017). When this is combined with SkipGram objective, an analysis of the objective shows that it asymptotically approximates the elementwise log of the seeded PageRank matrix. This fits into a broader research program to understand asymptotics of these sampling and fitting ideas (Levy and Goldberg 2014; Qiu et al. 2018; Chanpuriya and Musco 2020).

More directly relevant to our work, the research team behind the FREDE method (Tsitsulin et al. 2021) proposed to quickly compute an approximate embedding by randomly sampling PageRank vectors and using the SVD of the elementwise log of these vectors as the embedding. This is related to another method by Yin and Wei (2019). The FREDE procedure is exactly what we do, except the goal is a large dimensional embedding that might be used for graph learning tasks instead of the small dimensional embedding that we use here. In the context of Tsitsulin et al. (2021), our work establishes a new relationship between their methods and spectral clustering in the large α limit. Followup research seeks to accelerate these methods using sparse seeded PageRank and hashing (Postavaru et al. 2021).

Beyond the specific use of log PageRank, PageRank or diffusion based techniques have previously been used for learning graph embedding (or clustering) (Donnat et al. 2018; Klicpera et al. 2019; Yang et al. 2020; Takai et al. 2020; Liu et al. 2021; Carletti et al. 2020) where the personalized PageRank vector based on a set of nodes, called the **seed set** is used to focus regions and avoid unrolling neural networks over the entire graph.

9 Conclusion and future research

The key finding of this paper is that the elementwise log of a matrix of seeded PageRank vector approximates the spectral embedding of the Laplacian in degree regular graphs (Sect. 5). The methodology easily transfers to new scenarios such as hypergraphs given a PageRank-like primitive. This greatly simplifies the scenario compared with non-linear spectral methods on hypergraphs (Tudisco et al. 2021a, b; Tudisco and Higham 2021; Nguyen et al. 2017).

We believe our framework offers a successful technique for structural embedding and opens up some nontrivial research problems. Our code to compute the embeddings for these examples is available: https://github.com/dishashur/log-pagerank. We

believe this work lays the foundation for a reliable structural representation and the generalizability of this technique offers ample ground for new results.

The idea of customizing embeddings is highly relevant to the ongoing use of graph embeddings for ML algorithms. Customized embeddings quickly emerge from our framework by customizing the PageRank vectors – either by more sparsity or by customizing their seeding behavior.

Looking towards unsolved problems, the result on PageRank here seems a special case of a more general result about graph diffusions that scale the eigenvectors of the Laplacian. The heat kernel (Chung 2007) is another method, as well as a more general setting of arbitrary diffusion functions or polynomials (Kloster 2016) or learned diffusions (Jiang et al. 2017). Consequently, one future research direction includes studying the requirements on a diffusion function in order to guarantee this asymptotic limit. Another direction is to weaken the requirements on degree-regular graphs in the theory.

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Amburg, I., Veldt, N., Benson, A.R.: Clustering in graphs and hypergraphs with categorical edge labels. In: Proceedings of the Web Conference (2020)

Andersen, R., Chung, F., Lang, K.: Local graph partitioning using pagerank vectors. In: 2006 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06), pp. 475–486. IEEE (2006)

Bar-Yossef, Z., Mashiach, L.T.: Local approximation of PageRank and reverse PageRank. In: CIKM '08: Proceeding of the 17th ACM Conference on Information and Knowledge Management. ACM, New York, NY, USA, pp 279–288 (2008). https://doi.org/10.1145/1458082.1458122

Becchetti, L., Castillo, C., Donato, D., et al.: Link analysis for web spam detection. ACM Trans. Web **2**(1), 1–42 (2008). https://doi.org/10.1145/1326561.1326563

Benson, A.R., Abebe, R., Schaub, M.T., et al.: Simplicial closure and higher-order link prediction. Proc. Natl. Acad. Sci. (2018). https://doi.org/10.1073/pnas.1800683115

Bern, M., Mitchell, S., Ruppert, J.: Linear-size nonobtuse triangulation of polygons. In: Proceedings of the Tenth Annual Symposium on Computational Geometry, pp. 221–230. Association for Computing Machinery, New York, NY, USA, SCG '94 (1994). https://doi.org/10.1145/177424.177974,

Brin, S., Page, L.: The anatomy of a large-scale hypertextual web search engine. Comput. Netw. ISDN Syst. **30**(1), 107–117 (1998). https://doi.org/10.1016/S0169-7552(98)00110-X. Proceedings of the Seventh International World Wide Web Conference

Carletti, T., Battiston, F., Cencetti, G., et al.: Random walks on hypergraphs. Phys. Rev. E **101**(022), 308 (2020). https://doi.org/10.1103/PhysRevE.101.022308

Chanpuriya, S., Musco, C.: InfiniteWalk: Deep Network Embeddings as Laplacian Embeddings with a Nonlinearity, pp. 1325–1333. Association for Computing Machinery, New York (2020). https://doi.org/10.1145/3394486.3403185

Chung, F.R.L.: Spectral Graph Theory. American Mathematical Society, Providence (1992)

Chung, F.: The heat kernel as the pagerank of a graph. Proc. Natl. Acad. Sci. **104**(50):19,735–19,740 (2007). https://doi.org/10.1073/pnas.0708838104

Chung, F., Tsiatas, A., Xu, W.: Dirichlet pagerank and trust-based ranking algorithms. In: Frieze, A., Horn, P., Prałat, P. (eds.) Algorithms and Models for the Web Graph, pp. 103–114. Springer, Berlin (2011)

Constantine, P.G., Gleich, D.F.: Tall and skinny QR factorizations in MapReduce architectures. In: Proceedings of the Second International Workshop on MapReduce and Its Applications, pp. 43–50. ACM, New York, NY, USA, MapReduce '11 (2011). https://doi.org/10.1145/1996092.1996103

Constantine, P., Gleich, D.: Computing active subspaces with Monte Carlo (2014). arXiv preprint arXiv:1408.0545

Donnat, C., Zitnik, M., Hallac, D., et al.: Learning structural node embeddings via diffusion wavelets. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1320–1329. Association for Computing Machinery, New York, NY, USA, KDD'18 (2018). https://doi.org/10.1145/3219819.3220025,

Drineas, P., Ipsen, I.C.: Low-rank matrix approximations do not need a singular value gap. SIAM J. Matrix Anal. Appl. **40**(1), 299–319 (2019)

Fountoulakis, K., Liu, M., Gleich, D.F., et al.: Flow-based algorithms for improving clusters: A unifying framework, software, and performance (2020). arXiv:2004.09608

Frobenius, G.: Über matrizen aus nicht negativen elementen. Königliche Akademie der Wissenschaften Sitzungsber, Kön, pp. 456–477 (1912)

Gleich, D.F.: Models and algorithms for PageRank sensitivity (2009). Ph.D. thesis, Stanford University. http://www.stanford.edu/group/SOL/dissertations/pagerank-sensitivity-thesis-online.pdf

Gleich, D.F.: Pagerank beyond the web. SIAM Rev. **57**(3), 321–363 (2015). https://doi.org/10.1137/140976649

Gleich, D., Mahoney, M.: Anti-differentiating approximation algorithms: a case study with min-cuts, spectral, and flow. In: Xing, E.P., Jebara, T. (eds.) Proceedings of the 31st International Conference on Machine Learning, Proceedings of Machine Learning Research, vol. 32, pp. 1018–1025. PMLR, Bejing (2014). https://proceedings.mlr.press/v32/gleich14.html

Grover, A., Leskovec, J.: Node2vec: Scalable feature learning for networks. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 855–864. Association for Computing Machinery, New York, NY, USA, KDD '16 (2016). https://doi.org/10.1145/2939672.2939754

Hall, K.M.: An R-dimensional quadratic placement algorithm. Manag. Sci. **17**(3), 219–229 (1970)

Hammond, D.K., Vandergheynst, P., Gribonval, R.: Wavelets on graphs via spectral graph theory. Appl. Comput. Harmonic Anal. **30**(2), 129–150 (2011). https://doi.org/10.1016/j.acha.2010.04.005

Jiang, B., Kloster, K., Gleich, D.F., et al.: AptRank: an adaptive PageRank model for protein function prediction on bi-relational graphs. Bioinformatics **33**(12), 1829–1836 (2017). https://doi.org/10.1093/bioinformatics/btx029

Klicpera, J., Bojchevski, A., Günnemann, S.: Combining neural networks with personalized pagerank for classification on graphs. In: International Conference on Learning Representations (2019). https://openreview.net/forum?id=H1gL-2A9Ym

Kloster, K.: Graph diffusions and matrix functions: fast algorithms and localization results. Ph.D. thesis, Purdue University (2016). https://docs.lib.purdue.edu/open_access_dissertations/1404/

Koren, Y.: On spectral graph drawing. In: Warnow, T., Zhu, B. (eds.) Computing and Combinatorics, pp. 496–508. Springer, Berlin (2003)

Lang, K.: (2005) Fixing two weaknesses of the spectral method. In: Advances in Neural Information Processing Systems 18

Langville, A.N., Meyer, C.D.: Google's PageRank and Beyond: The Science of Search Engine Rankings. Princeton University Press, Princeton (2006)

Levy, O., Goldberg, Y.: (2014) Neural word embedding as implicit matrix factorization. In: Advances in Neural Information Processing Systems 27

Liu, M., Veldt, N., Song, H., et al.: Strongly local hypergraph diffusions for clustering and semi-supervised learning. In: Leskovec, J., Grobelnik, M., Najork, M., et al. (eds) WWW '21: The Web Conference 2021, Virtual Event/Ljubljana, Slovenia, April 19–23, 2021, pp 2092–2103. ACM/IW3C2 (2021). https://doi.org/10.1145/3442381.3449887,

Mahoney, M.W., Orecchia, L., Vishnoi, N.K.: A local spectral method for graphs: with applications to improving graph partitions and exploring data graphs locally. J. Mach. Learn. Res. **13**(1), 2339–2365 (2012)

Nguyen, Q., Tudisco, F., Gautier, A., et al.: An efficient multilinear optimization framework for hypergraph matching. IEEE Trans. Pattern Anal. Mach. Intell. **39**(6), 1054–1075 (2017)

Perozzi, B., Al-Rfou, R., Skiena, S.: Deepwalk: online learning of social representations. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Association for Computing Machinery, New York, NY, USA, KDD '14, pp. 701–710 (2014) . https://doi.org/10.1145/2623330.2623732

Perron, O.: Zur theorie der matrices. Math. Ann. **64**(2), 248–263 (1907)

Postavaru, S., Tsitsulin, A., de Almeida, F.M.G., et al.: Instantembedding: Efficient local node representations (2021). https://openreview.net/forum?id=4vDf4Qtodh

Pothen, A., Simon, H.D., Liou, K.P.: Partitioning sparse matrices with eigenvectors of graphs. SIAM J. Matrix Anal. Appl. **11**, 430–452 (1990). https://doi.org/10.1137/0611030

Qiu, J., Dong, Y., Ma, H., et al.: Network embedding as matrix factorization: Unifying deepwalk, line, pte, and node2vec. In: Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining. Association for Computing Machinery, New York, NY, USA, WSDM '18, pp. 459–467 (2018). https://doi.org/10.1145/3159652.3159706

Serra-Capizzano, S.: Jordan canonical form of the google matrix: a potential contribution to the pagerank computation. SIAM J. Matrix Anal. Appl. **27**(2), 305–312 (2005). https://doi.org/10.1137/S0895479804441407

Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **22**(8), 888–905 (2000). https://doi.org/10.1109/34.868688

Stehlé, J., Voirin, N., Barrat, A., et al.: High-resolution measurements of face-to-face contact patterns in a primary school. PLoS ONE **6**(8), e23,176 (2011). https://doi.org/10.1371/journal.pone.0023176,

Stewart, G.W.: Error and perturbation bounds for subspaces associated with certain eigenvalue problems. SIAM Rev. **15**(4), 727–764 (1973)

Takai, Y., Miyauchi, A., Ikeda, M., et al.: Hypergraph Clustering Based on PageRank, Association for Computing Machinery, New York, NY, USA, p 1970–1978 (2020). https://doi.org/10.1145/3394486.3403248

Tang, J., Qu, M., Mei, Q.: Pte: Predictive text embedding through large-scale heterogeneous text networks. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Association for Computing Machinery, New York, NY, USA, KDD '15, pp. 1165–1174 (2015a). https://doi.org/10.1145/2783258.2783307

Tang, J., Qu, M., Wang, M., et al.: Line: Large-scale information network embedding. In: Proceedings of the 24th International Conference on World Wide Web. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, WWW '15, pp. 1067–1077 (2015b). https://doi.org/10.1145/2736277.2741093

Tong, H., Faloutsos, C., Pan, J.Y.: Fast random walk with restart and its applications. In: Sixth International Conference on Data Mining (ICDM'06). IEEE, pp. 613–622 (2006)

Tropp, J.A.: User-friendly tail bounds for sums of random matrices. Found. Comput. Math. **12**(4), 389–434 (2012)

Tsitsulin, A., Munkhoeva, M., Mottin, D., et al.: FREDE: anytime graph embeddings. Proc. VLDB Endow. **14**(6), 1102–1110 (2021). https://doi.org/10.14778/3447689.3447713

Tudisco, F., Higham, D.J.: Node and edge nonlinear eigenvector centrality for hypergraphs (2021). arXiv:2101.06215

Tudisco, F., Benson, A.R., Prokopchik, K.: Nonlinear higher-order label spreading. In: Proceedings of the Web Conference 2021. Association for Computing Machinery, New York, NY, USA, WWW '21, pp. 2402–2413 (2021a). https://doi.org/10.1145/3442381.3450035

Tudisco, F., Prokopchik, K., Benson, A.R.: A nonlinear diffusion method for semi-supervised learning on hypergraphs (2021b). arXiv:2103.14867

Veldt, N., Benson, A.R., Kleinberg, J.: Minimizing localized ratio cut objectives in hypergraphs. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM Press (2020a)

Veldt, N., Benson, A.R., Kleinberg, J.: Minimizing localized ratio cut objectives in hypergraphs. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 1708–1718 (2020b)

Yang, R., Shi, J., Xiao, X., et al.: Homogeneous network embedding for massive graphs via reweighted personalized pagerank. Proc VLDB Endow **13**(5), 670–683 (2020). https://doi.org/10.14778/3377369.3377376

Yin, Y., Wei, Z.: Scalable graph embeddings via sparse transpose proximities. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Association for Computing Machinery, New York, NY, USA, KDD '19, pp. 1429–1437 (2019). https://doi.org/10.1145/3292500.3330860

Zhou, C., Liu, Y., Liu, X., et al.: Scalable graph embedding for asymmetric proximity. In: Proceedings of the AAAI Conference on Artificial Intelligence 31(1). https://doi.org/10.1609/aaai.v31i1.10878. https://ojs.aaai.org/index.php/AAAI/article/view/10878 (2017)