

PageRank and random walks on graphs

Fan Chung and Wenbo Zhao

University of California, San Diego
La Jolla, CA 92093, US
{fan, pedu, w3zhao}@ucsd.edu

Dedicated to Lovász on the occasion of his sixtieth birthday.

Abstract. We examine the relationship between PageRank and several invariants occurring in the study of random walks and electrical networks. We consider a generalized version of hitting time and effective resistance with an additional parameter which controls the ‘speed’ of diffusion. We will establish their connection with PageRank. Through these connections, a combinatorial interpretation of PageRank is given in terms of rooted spanning forests by using a generalized version of the matrix-tree theorem. Using PageRank, we will illustrate that the generalized hitting time leads to finding sparse cuts and efficient approximation algorithms for PageRank can be used for approximating hitting time and effective resistance.

1 Introduction

The notion of PageRank, first introduced by Brin and Page [2], forms the basis for their Web search algorithms. Although the original version of PageRank was used for the Webgraph (with all the webpages as vertices and hyperlinks as edges), PageRank is well defined for any given graph and is quite effective for capturing various relations among vertices of graphs. In this paper, we will investigate several implications of PageRank for a given graph.

To start with, we give the graph-theoretical definition of PageRank. Roughly speaking, PageRank is a way to organize random walk of various lengths. Instead of having to determine the number of steps a random walk is taking, PageRank uses a positive real value α , where $\alpha \in [0, 1)$ to control the “diffusion” of a combination of random walks.

The original definition for PageRank was to assign a value to each vertex (Webpage), denoting the “importance” of a vertex under two assumptions: For some fixed probability α , a surfer at a Webpage jumps to a random Webpage with probability α and goes to a linked Webpage with probability $1 - \alpha$. The importance of a Webpage v is the expected sum of the importance of all the Webpages u that precede v .

In this paper, we will use a more general version of PageRank, called *personalized PageRank*, introduced by Jeh and Widom [7] (also see Haveliwala [6]). The personalized PageRank $\text{pr}_\alpha(s)$ depends on two parameters, the *jumping constant*

α and a *seed* s . A seed can be viewed as a vertex or a probability distribution on vertices. The original definition of PageRank is the special case where the seed is the uniform distribution.

To define the PageRank for a connected graph G , we consider random walks on G with transition probability matrix P and the lazy random walk on G , denoted by $Z = (I + P)/2$. In this paper, all vectors are regarded as row vectors unless stated otherwise.

The personalized PageRank vector $\text{pr}_\alpha(s)$ with a jumping constant α and a seed vector s is defined to be the unique solution of the linear system

$$\text{pr}_\alpha(s) = \alpha s + (1 - \alpha)\text{pr}_\alpha(s)Z. \quad (1)$$

An alternate but equivalent definition for pr_α is an expression of a geometric sum of random walks:

$$\text{pr}_\alpha(s) = \alpha s \sum_{k=0}^{\infty} (1 - \alpha)^k Z^k.$$

In addition to the practical applications of Websearch algorithms, PageRank has numerous connections to various graph invariants. For example, PageRank can be used to find cuts with a certain isoperimetric guarantee, similar to the Cheeger inequalities. It was shown in [1] that for an arbitrarily chosen vertex u , if we arrange vertices in a row using PageRank $\text{pr}_\alpha(u)$, one of the cuts which consists of vertices in initial segments has Cheeger ratio optimal up to a quadratic factor with high probability. The performance guarantee is quite similar to that given by the spectral partitioning algorithm using eigenvectors. However, the advantages of using PageRank are multifold. We can choose an appropriate α to specify the approximate size of the part that we wish to cut and thus PageRank leads to so-called *local* algorithms. Furthermore, there are effective algorithms for computing approximate PageRank with finite support with size depending only on the error bound and the desired size of the smaller separated part [1].

In this paper, we explore the relationship of PageRank and various graph invariants occurring in random walks and electrical networks. In the spirit of PageRank, we consider a generalized version of the hitting time and effective resistance with an additional parameter α . In a way, these generalized invariants provide a quantitative ranking of edges, indicating how important an edge is while allowing the choice of α . In Section 2, we will define the Laplacian and the discrete Green's function as well as their connection to PageRank. In Section 3, we consider electrical networks and the generalized versions of hitting time and the effective resistance. In Section 4, we will give several matrix-forest theorems, which generalize the classical Matrix-Tree Theorem [11]. In Section 5, we will derive a combinatorial interpretation of PageRank in terms of spanning forests in the graph. In Section 6, we consider some useful properties of the generalized hitting time in connection of identifying sparse cuts. In Section 7, we use PageRank to estimate the effective resistance.

2 Laplacian, the Green's function and PageRank

We consider a connected weighted undirected graph $G = (V, E, w)$. Suppose G has vertex set V , edge set E , edge weight $w_{u,v} \geq 0$ and $|V| = n, |E| = m$. A typical random walk is determined by the transition probabilities $P(u, v) = w_{uv}/d_u$ where the degree d_u of u is the sum $\sum_{v:\{u,v\} \in E} w_{u,v}$. The volume of a subset $S \subseteq V$, denoted by $\text{vol}(S)$, is the sum of degrees of vertices in S . In particular, the volume of G , denoted by $\text{vol}(G)$, is equal to $\text{vol}(V)$. For the special case of $w_{u,v} = 1$ for all $\{u, v\} \in E$, we have $\text{vol}(G) = 2m$.

Let A denote the weighted adjacency matrix with entries $A(u, v) = w_{u,v}$ and D denote the diagonal degree matrix. Then the transition probability matrix P is equal to $D^{-1}A$ and for any initial distribution f , the distribution of the random walk after k steps is fP^k . For undefined terminology, the reader is referred to the excellent survey of Lovász [13]

The combinatorial Laplacian of G is defined by $L = D - A$. If we orient the edges of G in an arbitrary but fixed way, we can write its Laplacian as

$$L = B^T W B, \quad (2)$$

where B is the signed edge-vertex incidence matrix, given by

$$B(e, v) = \begin{cases} 1 & \text{if } v \text{ is } e\text{'s head} \\ -1 & \text{if } v \text{ is } e\text{'s tail} \\ 0 & \text{otherwise} \end{cases}$$

and W is the diagonal matrix with $W(e, e) = w_e$. The normalized Laplacian of G is defined to be $\mathcal{L} = D^{-1/2} L D^{-1/2}$ and we can write

$$\mathcal{L} = S^T W S$$

where $S = B D^{-1/2}$.

Since \mathcal{L} is symmetric and we can express \mathcal{L} by

$$\mathcal{L} = \sum_{i=0}^{n-1} \lambda_i \phi_i^T \phi_i = \sum_{i=1}^{n-1} \lambda_i \phi_i^T \phi_i,$$

where $\lambda_0 = 0$ and $0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_{n-1} \leq 2$ are the nonzero eigenvalues of \mathcal{L} and $\phi_0, \dots, \phi_{n-1}$ form a corresponding orthonormal basis of eigenvectors. The fact of $\lambda_1 > 0$ follows from the connectivity of G . The eigenvalue λ_1 is intimately related to the rate of convergence of random walks. The reader is referred to [4] for numerous properties concerning eigenvalues of the normalized Laplacian. Although the combinatorial Laplacian is useful for various flow problems in the study of electrical networks, the spectrum of combinatorial Laplacian is not effective (except for almost regular graphs) for applications requiring isoperimetric properties.

Denote the β -normalized Laplacian \mathcal{L}_β by $\beta I + \mathcal{L}$. Then we may write $\mathcal{L}_\beta = S'^T W_\beta S'$ where we define S' and W_β as follows:

$$S' = \begin{bmatrix} I \\ S \end{bmatrix}_{(n+m) \times n} \quad B' = \begin{bmatrix} D^{1/2} \\ B \end{bmatrix}_{(n+m) \times n} \quad \text{and } W_\beta = \begin{bmatrix} \beta I & 0 \\ 0 & W \end{bmatrix}_{(n+m) \times (n+m)}.$$

For simplicity, we index the columns of S' and the columns of W_β by $V \cup E$ where the first n columns are indexed by V and the last m columns are indexed by E . The rows of W_β are indexed in the same way. We define

$$\mathcal{L}_\beta = D^{-1/2} B'^T W_\beta B' D^{-1/2} = S'^T W_\beta S'.$$

Green's functions were first introduced in a celebrated essay by George Green [8] in 1828. Since then, the concept of Green's functions has been used in a wide range of areas, especially in the study of partial differential equations and quantum field theory. The discrete analog of Green's functions, which are associated with the normalized Laplacian of graphs, were first introduced in a 2002 paper [5] in connection with the study of Dirichlet eigenvalues with boundary conditions. The Green's function \mathcal{G} denotes the symmetric matrix satisfying

$$f \mathcal{L} \mathcal{G} = f \mathcal{G} \mathcal{L} = f.$$

for all vectors f which are orthogonal to the eigenvector $\phi_0 = \mathbf{1} D^{1/2} / \sqrt{\text{vol}(G)}$ where $\mathbf{1}$ denotes the all 1's vector. The Green's function \mathcal{G} has the following form:

$$\mathcal{G} = \sum_{i=1}^{n-1} \frac{1}{\lambda_i} \phi_i^T \phi_i. \quad (3)$$

The following modified Green's function \mathcal{G}_β was also used in [5]. For $\beta \in \mathbb{R}^+$, let Green's function \mathcal{G}_β denote the symmetric matrix satisfying

$$\mathcal{L}_\beta \mathcal{G}_\beta = I. \quad (4)$$

Clearly, we have

$$\mathcal{G}_\beta = \sum_{i=0}^{n-1} \frac{1}{\lambda_i + \beta} \phi_i^T \phi_i. \quad (5)$$

By comparing with the recurrence of the PageRank in (1), we remark that the discrete Green's function is basically a symmetric form of the PageRank. Namely, we can write

$$\frac{\text{pr}_\alpha(s)}{\beta} = s D^{-1/2} \mathcal{G}_\beta D^{1/2} \quad (6)$$

where

$$\beta = \frac{2\alpha}{1-\alpha}. \quad (7)$$

3 PageRank, the hitting time and the effective resistance in electrical networks

A connected weighted undirected graph $G = (V, E, w)$ can be viewed as an electrical network, where the edge weight w_e represents the *conductance* of e .

The *effective resistance* $R(u, v)$ between two vertices u and v is defined as the voltage potential difference induced between them when a unit current is injected at u and extracted at v . The effective resistance can be characterized by the combinatorial Laplacian of the graph (see [11]). Since we wish to establish the connection of the effective resistance with PageRank, we will consider the normalized Laplacian instead.

Suppose we are given the injected current function $i_V : V \rightarrow \mathbb{R}$. The induced current i_E on the edges satisfies the property that the sum of all induced current on edges entering v is equal to $i_V(v)$, as asserted by *Kirchoff's current law*:

$$i_V = i_E B.$$

For any function $f : V \rightarrow \mathbb{R}$, we can regard f as a *voltage potential* function in the following sense: The induced flow for the edge from u to v is the product of $f(u) - f(v)$ and the conductance of the edge, according to *Ohm's law*, and can be expressed by

$$i_E = f_V B^T W.$$

We can write

$$i_V D^{-1/2} = (f_V B^T W) B D^{-1/2} = f_V D^{1/2} \mathcal{L}.$$

Suppose we only consider the voltage potential function f satisfying $\sum_v f(v) = 0$. By using the definition of Green's function, we have

$$i_V D^{-1/2} \mathcal{G} D^{-1/2} = f_V.$$

Suppose we inject a unit current to vertex u and extract a unit current from v , i.e., $i_V = \chi_v - \chi_u$ where χ_u is the characteristic function with $\chi_u(x) = 1$ if $x = u$ and 0 otherwise. Thus, the effective resistance between vertices u and v can be written as

$$\begin{aligned} R(u, v) &= f_V (\chi_v - \chi_u)^T \\ &= i_V D^{-1/2} \mathcal{G} D^{-1/2} (\chi_v - \chi_u)^T \\ &= (\chi_v - \chi_u) D^{-1/2} \mathcal{G} D^{-1/2} (\chi_v - \chi_u)^T. \end{aligned} \tag{8}$$

The effective resistance is closely associated with the hitting time and commute time for a random walk on G . The hitting time $H(u, v)$ is the expected number of steps of a random walk starting from u until it first visit v . The commute time $C(u, v)$ of u and v is the expected number of steps in a random walk starting at u , before vertex v is visited, and the vertex u is reached again. The commute time can be expressed as the sum

$$C(u, v) = H(u, v) + H(v, u).$$

It has been shown [12] (also see [3]) that the commute time $C(u, v)$ satisfies

$$C(u, v) = \frac{R(u, v)}{\text{vol}(G)}.$$

For two distinct vertices u and v , the hitting time $H(u, v)$ satisfies the following equation:

$$H(u, v) = 1 + \frac{1}{d_u} \sum_{w:(w,u) \in E} H(w, v) \quad (9)$$

and $H(u, u) = 0$. Here we will express the hitting time in terms of the Green's function.

Lemma 1. *For all $u, v \in V$,*

$$\frac{H(u, v)}{\text{vol}(G)} = (\chi_v - \chi_u) D^{-1/2} \mathcal{G} D^{-1/2} \chi_v^T.$$

Proof. Clearly, $H(v, v) = 0$ for all $v \in V$. We use the equation in (9). Consider

$$\begin{aligned} & \frac{1}{\text{vol}(G)} + \frac{1}{d_u} \sum_{w:(w,u) \in E} (\chi_v - \chi_u) D^{-1/2} \mathcal{G} D^{-1/2} \chi_v^T \\ &= \frac{1}{\text{vol}(G)} + \chi_v D^{-1/2} \mathcal{G} D^{-1/2} \chi_v^T - \chi_u D^{-1} A D^{-1/2} \mathcal{G} D^{-1/2} \chi_v^T \\ &= \frac{1}{\text{vol}(G)} + \chi_v D^{-1/2} \mathcal{G} D^{-1/2} \chi_v^T - \chi_u D^{-1/2} (I - \mathcal{L}) \mathcal{G} D^{-1/2} \chi_v^T \\ &= (\chi_v - \chi_u) D^{-1/2} \mathcal{G} D^{-1/2} \chi_v^T + \chi_u D^{-1/2} \left(\frac{1}{\text{vol}(G)} D^{1/2} J D^{1/2} + \mathcal{L} \mathcal{G} \right) D^{-1/2} \chi_v^T \\ &= (\chi_v - \chi_u) D^{-1/2} \mathcal{G} D^{-1/2} \chi_v^T + \chi_u D^{-1/2} \left(\frac{1}{\text{vol}(G)} D^{1/2} J D^{1/2} + I - u_0^T u_0 \right) D^{-1/2} \chi_v^T \\ &= (\chi_v - \chi_u) D^{-1/2} \mathcal{G} D^{-1/2} \chi_v^T + \chi_u D^{-1} \chi_v^T \\ &= (\chi_v - \chi_u) D^{-1/2} \mathcal{G} D^{-1/2} \chi_v^T \end{aligned}$$

□

We define

$$h(u, v) = \frac{H(u, v)}{\text{vol}(G)}.$$

Now, we define a generalized hitting time with an additional parameter $\alpha > 0$.

$$h_\alpha(u, v) = \frac{[\text{pr}_\alpha(\chi_v)](v)}{d_v} - \frac{[\text{pr}_\alpha(\chi_v)](u)}{d_u}. \quad (10)$$

We define the generalized effective resistance $R_\alpha(u, v)$ as follows:

$$R_\alpha(u, v) = h_\alpha(u, v) + h_\alpha(v, u).$$

It is easy to check that $R_\alpha(u, v)$ satisfies

$$R_\alpha(u, v) = \beta (\chi_v - \chi_u) D^{-1/2} \mathcal{G}_\beta D^{-1/2} (\chi_v - \chi_u)^T$$

where β satisfies (7). By using the connection with PageRank in (6), we can write

$$R_\alpha(u, v) = \frac{[\text{pr}_\alpha(\chi_u)](u)}{d_u} - \frac{[\text{pr}_\alpha(\chi_u)](v)}{d_v} + \frac{[\text{pr}_\alpha(\chi_v)](v)}{d_v} - \frac{[\text{pr}_\alpha(\chi_v)](u)}{d_u}. \quad (11)$$

4 Several matrix-forest theorems

It is well-known that in a simple graph G for an edge joining u and v , the effective resistance $R(u, v)$ is proportional to the number of spanning trees containing the edge $\{u, v\}$. Later on we will show that the generalized effective resistance $R_\alpha(u, v)$ is proportional to a combination of spanning forests of certain types. A *forest* is a graph containing no cycle. A k -rooted forest is a forest with k connected components where each of the connected components contains a special vertex that we call a root. A k -rooted spanning forest is a k -rooted forest containing all vertices as vertices. It has exactly $n - k$ edges. The special case for $k=1$ is a rooted spanning tree.

The weight of a forest F is defined to be the product of all weights of edges in F . We define the weight $w(F)$ of an unrooted forest as follows:

$$w(F) = \prod_{e \in F} w(e).$$

For example, if F is a rooted tree in a simple graph, then $w(F) = 1$. For a rooted tree F , we denote by F_* the set of all vertices in F which are not roots. The weight $w^*(F)$ of a rooted forest is defined by

$$w^*(F) = \frac{\prod_{e \in F} w(e)}{\prod_{v \in F_*} d_v}.$$

First, we prove several useful facts along the same lines as the Matrix-Tree Theorem [11]

Theorem 1. *In a graph $G = (V, E, w)$, for $\beta > 0$, the determinant of \mathcal{L}_β is related to rooted spanning forests as follows:*

$$\det \mathcal{L}_\beta = \sum_{k=1}^n \beta^k \sum_{F \in \mathcal{F}_k} w^*(F)$$

where \mathcal{F}_k denotes the family of all k -rooted spanning forests in G .

Proof. For a matrix M and subsets X, Y of indices of rows and columns, we denote a submatrix $M_{X, Y}$ of M by restricting rows and columns of M to X and Y . We consider the determinant of \mathcal{L}_β as follows:

$$\begin{aligned} \det \mathcal{L}_\beta &= \det S'^T W_\beta S' \\ &= \sum_{\substack{F \subseteq V \cup E \\ |F|=n}} (\det S'_{F, V}) (\det W_{F, F}) (\det S'_{F, V}) \end{aligned}$$

by the Cauchy-Binet Theorem. We note that $\det S'_{F,V} = 0$ if F contains a cycle. In the other direction, if F does not contain a cycle, F is a spanning forest and the vertices in $F \cap V$ are the k roots of the spanning forest. Therefore, we have

$$\det \mathcal{L}_\beta = \sum_{k=1}^n \sum_{\substack{F \\ k\text{-rooted spanning forest} \\ \text{with roots in } F \cap V}} (\det S'_{F,V})^2 \det(W_\beta)_{F,F}.$$

Note that for a k -rooted spanning forest F , we have

$$\det S'_{F,V} = \pm \frac{\beta^k}{\sqrt{\prod_v d_v}} \prod_{u \in F \cap V} \sqrt{d_u}.$$

Thus we have

$$\begin{aligned} \det \mathcal{L}_v &= \frac{1}{\prod_v d_v} \sum_{k=1}^n \beta^k \sum_{\substack{F \\ k\text{-rooted spanning forests}}} \prod_{u \in F \cap V} d_u \prod_{e \in F} w_e \\ &= \sum_{k=1}^n \beta^k \sum_{F \in \mathcal{F}_k} w^*(F) \end{aligned}$$

as claimed. \square

Theorem 2. *In a graph $G = (V, E, w)$, for $\beta > 0$ and a vertex v in G , the determinant of the principle minor $\mathcal{L}_\beta^{(v)}$, obtained by excluding the row and column associated with v satisfies :*

$$\det \mathcal{L}_\beta^{(v)} = \sum_{k=1}^n \beta^{k-1} \sum_{F \in \mathcal{F}_{k,v}} w^*(F)$$

where $\mathcal{F}_{k,v}$ denotes the family of k -rooted trees having v as one of the roots.

Proof. Let V' denote $V \setminus \{v\}$. We can write the determinant of $\mathcal{L}_\beta^{(v)}$ as follows:

$$\begin{aligned} \det \mathcal{L}_\beta^{(v)} &= \det \left((S'_{V' \cup E, V'})^T W_\beta S'_{V' \cup E, V'} \right) \\ &= \sum_{\substack{F \subseteq V' \cup E \\ |F \cap V'| = k-1 \\ |F| = n-1}} (\det S'_{F, V'})^2 \det(W_\beta)_{F,F} \\ &= \sum_{k=1}^n \sum_{\substack{F \\ k\text{-rooted spanning forest} \\ \text{with roots in } F \cap V' \text{ and } v}} (\det S'_{F, V'})^2 \det(W_\beta)_{F,F} \end{aligned}$$

since if the forest F satisfies $\det S'_{F, V'} \neq 0$, then $F \subseteq V' \cup E$ and F forms a k -rooted spanning forest with v as a root. Thus,

$$\det S'_{F, V'} = \pm \frac{\beta^{k-1}}{\sqrt{\prod_{u \neq v} d_u}} \prod_{u \in F \cap V} d_u,$$

and we have

$$\begin{aligned} \det \mathcal{L}_\beta^{(v)} &= \frac{1}{\prod_{u \neq v} d_u} \sum_{k=1}^n \beta^{k-1} \sum_{\substack{F \\ k\text{-rooted spanning forests} \\ v \text{ is a root}}} \prod_{u \in F \cap V} d_u \prod_{e \in F} w_e \\ &= \sum_{k=1}^n \beta^{k-1} \sum_{F \in \mathcal{F}_{k,v}} w^*(F) \end{aligned}$$

as desired. \square

Theorem 3. *In a graph $G = (V, E, w)$, for $\beta > 0$ and two distinct vertices u, v in G , the determinant of the minor $\mathcal{L}_\beta^{(u,v)}$, obtained by excluding the row associated with u and column associated with v satisfies :*

$$\det \mathcal{L}_\beta^{(u,v)} = \sqrt{\frac{d_v}{d_u}} \sum_{k=1}^n \beta^{k-1} \sum_{F \in \mathcal{F}_{k,u,v}} w^*(F)$$

where $\mathcal{F}_{k,u,v}$ denotes the family of k -rooted spanning forests in which u and v are in the same connected component and u is a root.

Proof. We denote $V_1 = V \setminus \{u\}$ and $V_2 = V \setminus \{v\}$. We consider the determinant of $\mathcal{L}_\beta^{(u,v)}$ as follows:

$$\begin{aligned} \det \mathcal{L}_\beta^{(u,v)} &= \sigma(u, v) \det \left((S'_{V_1 \cup E, V})^T W_\beta S'_{V_2 \cup E, V} \right) \\ &= \sigma(u, v) \sum_{k=1}^n \sum_{\substack{F \subseteq V \cup E \\ |F \cap V_1| = k-1 \\ |F| = n-1}} (\det S'_{F, V_1}) (\det S'_{F, V_2}) \det (W_\beta)_{F, F} \\ &= \sigma(u, v) \sum_{k=1}^n \sum_{\substack{F \subseteq V \cup E \\ |F \cap V''| = k-1 \\ |F| = n-1}} (\det S'_{F, V_1}) (\det S'_{F, V_2}) \det (W_\beta)_{F, F} \end{aligned}$$

since if u or v is in F , one of the determinants in the product of the preceding equality becomes zero. Here $\sigma(u, v)$ denotes $(-1)^{i+j}$ if u is the index of the i th row and v is the index of the j th column of \mathcal{L} . If $\det S'_{F, V_1}$ is non-zero, the edges in F form a spanning tree with roots $\{u\} \cup (F \cap V'')$ where $V'' = V \setminus \{u, v\}$. If $\det S'_{F, V_2}$ is non-zero, then the edges in F form a spanning tree with roots $\{v\} \cup (F \cap V'')$. Therefore if the product is nonzero, vertices u and v are in the same connected component in the spanning forest formed by the edges in F . Suppose $F \subseteq V'' \cup E$ forms a rooted spanning forest with k connected components, one of which contains both u and v . Thus we have

$$\det \mathcal{L}_\beta^{(u,v)} = \frac{\sigma(u, v)}{\sqrt{d_u d_v} \prod_{x \neq u, v} d_x} \sum_{k=1}^n \sum_{\substack{F \subseteq V'' \cup E \\ |F \cap V''| = k-1 \\ |F| = n-1}} (\det B'_{F, V_1}) (\det B'_{F, V_2}) \det (W_\beta)_{F, F}$$

We will prove the following Claim later.

Claim:

$$\det B'_{F,V_2} = \sigma(u, v) \det B'_{F,V_1}.$$

From the Claim, we have

$$\begin{aligned} \det \mathcal{L}_\beta^{(u,v)} &= \frac{1}{\sqrt{d_u d_v} \prod_{x \neq u,v} d_x} \sum_{k=1}^n \sum_{\substack{F \\ k\text{-rooted spanning forest} \\ u \text{ is a root} \\ u, v \text{ are in the same c. c.}}} (\det B'_{F,V_1})^2 \det(W_\beta)_{F,F} \\ &= \frac{1}{\sqrt{d_u d_v} \prod_{x \neq u,v} d_x} \sum_{k=1}^n \sum_{F \in \mathcal{F}_{k,u,v}} (\det B'_{F,V_1})^2 \det(W_\beta)_{F,F} \\ &= \frac{1}{\sqrt{d_u d_v} \prod_{x \neq u,v} d_x} \sum_{k=1}^n \sum_{F \in \mathcal{F}_{k,u,v}} \beta^{k-1} \prod_{e \in F \cap E} w_e \prod_{x \in F \cap V} d_x \\ &= \frac{1}{\sqrt{d_u d_v}} \sum_{k=1}^n \beta^{k-1} \sum_{F \in \mathcal{F}_{k,u,v}} \frac{\prod_{e \in F} w_e}{\prod_{x \in F^* \setminus \{v\}} d_x} \\ &= \sqrt{\frac{d_v}{d_u}} \sum_{k=1}^n \beta^{k-1} \sum_{F \in \mathcal{F}_{k,u,v}} w^*(F). \end{aligned}$$

It remains to prove the Claim.

Proof of the Claim:

We prove the claim in three steps:

Case 1: First we see that the Claim holds if F forms a path with vertices $u = u_1, u_2, \dots, u_{k+1} = v$ and edges $e_i = \{u_i, u_{i+1}\}$ in F where u_i is the index for the i th row of \mathcal{L} , and e_j is the index of the j th column in B .

Case 2: Since u and v are in the same connected component of F , let $u = v_1, v_2, \dots, v_t = v$ denote the path joining u and v in F . If u_i is the index for the i th row of \mathcal{L} , and e_j is the index of the j th column in B for $1 \leq i \leq t+1$ and $1 \leq j \leq t$, then the claim holds in the same way as in Case 1.

Case 3: Now we consider the general case. We follow the notation in Case 2 concerning the paths joining u and v in F . We denote by U the permutation on V that moves v_i to the i th place for $i = 1, \dots, k+1$ and denote by U' the permutation on E that moves $e'_j = \{v_j, v_{j+1}\}$ to the j th place for $j = 1, \dots, k$. We consider

$$\begin{aligned} \sigma(u, v) \det ((B'_{F,V_1})^T B'_{F,V_2}) &= \sigma(U(u), U(v)) \det (U B'_{F,V_1}{}^T U'^T U' B'_{F,V_2} V) \\ &= \sigma(U(u), U(v)) \det (U'^T B'_{F,V_1} U) \det (U' B'_{F,V_2} V) \end{aligned}$$

We have reduced Case 3 to Case 2. Thus the above expression is equal to 1 and the Claim is proved.

The proof of Theorem 3 is completed. \square

5 PageRank and other invariants in terms of rooted spanning forests

Using the matrix-forest theorems in the preceding section, we can establish the connection between the discrete Green's function and rooted spanning forests. We consider the discrete Green's function \mathcal{G}_β for $\beta > 0$. For a fixed vertex v , we consider

$$\mathcal{L}_\beta f = \chi_v. \quad (12)$$

Clearly, $f = \mathcal{G}_\beta \chi_v$ is a solution to the above equation. We can view (12) as a linear system with variables $f(u)$ for u in V . By Cramer's rule, we have

$$f(v) = \mathcal{G}_\beta(v, v) = \frac{\det \mathcal{L}_\beta^{(v)}}{\det \mathcal{L}_\beta} \quad (13)$$

$$f(u) = \mathcal{G}_\beta(u, v) = \frac{\det \mathcal{L}_\beta^{(u,v)}}{\det \mathcal{L}_\beta}. \quad (14)$$

From equation (6) and Theorems 1, 2 and 3, we can express PageRank in terms of the combinatorial sums involving rooted spanning forests as follows:

Theorem 4. *For a vertex v in G , the PageRank pr_α satisfies*

$$[\text{pr}_\alpha(\chi_v)](v) = \frac{\sum_{k=1}^n \beta^k \sum_{F \in \mathcal{F}_{k,v}} w^*(F)}{\sum_{k=1}^n \beta^k \sum_{F \in \mathcal{F}_k} w^*(F)}$$

where $\beta = 2\alpha/(1-\alpha)$, \mathcal{F}_k denotes the family of all k -rooted spanning forests in G and $\mathcal{F}_{k,v}$ denotes the family of k -rooted trees having v as one of the roots.

Theorem 5. *For two distinct vertices u and v in G , the PageRank pr_α satisfies*

$$[\text{pr}_\alpha(\chi_u)](v) = \frac{d_v \sum_{k=1}^n \beta^k \sum_{F \in \mathcal{F}_{k,u,v}} w^*(F)}{d_u \sum_{k=1}^n \beta^k \sum_{F \in \mathcal{F}_k} w^*(F)}$$

where $\beta = 2\alpha/(1-\alpha)$ and $\mathcal{F}_{k,u,v}$ denotes the family of k -rooted spanning forests in which u and v are in the same connected component and u is a root.

As it turns out, the generalized hitting time can also be expressed in a similar form:

Theorem 6. *For an edge $e = \{u, v\}$ in G , the hitting time $h_\alpha(u, v)$ satisfies*

$$h_\alpha(u, v) = \frac{\beta \sum_{k=1}^{n-1} \beta^k \sum_{F \in \mathcal{F}_{k,e,v}} w^*(F)}{w_e \sum_{k=1}^n \beta^k \sum_{F \in \mathcal{F}_k} w^*(F)}$$

where $\beta = 2\alpha/(1-\alpha)$ and $\mathcal{F}_{k,e,v}$ denotes the family of k -rooted spanning forests containing e as an edge which is in the path from v to its root.

Proof. From (10), we have

$$\begin{aligned} h_\alpha(u, v) &= \frac{[\text{pr}_\alpha(\chi_v)](v)}{d_v} - \frac{[\text{pr}_\alpha(\chi_v)](u)}{d_u} \\ &= \frac{\sum_{k=1}^n \beta^k (\sum_{F \in \mathcal{F}_{k,v}} w^*(F) - \sum_{F \in \mathcal{F}_{k,v,u}} w^*(F))}{d_v \sum_{k=1}^n \beta^k \sum_{F \in \mathcal{F}_k} w^*(F)} \end{aligned}$$

Note that the set of $\mathcal{F}_{k,v} \setminus \mathcal{F}_{k,v,u}$ consists of k -rooted trees in which u and v are in different connected components and v is a root. For every rooted forest F in $\mathcal{F}_{k,v} \setminus \mathcal{F}_{k,v,u}$ we consider the rooted tree F' formed by adding the edge $e = \{u, v\}$ where v is no longer a root. The path from v to its root in F' contains e . Clearly, F' has $k-1$ connected components and $k-1$ roots. Let $\mathcal{F}_{k-1,e,v}$ denote the family of $k-1$ forests with the additional property that the path from v to its root contains e . Thus F' is in $\mathcal{F}_{k-1,e,v}$. We note

$$w^*(F') = \frac{w_e}{d_v} w^*(F)$$

since v is no longer a root. This is a bijection from $\mathcal{F}_{k,v} \setminus \mathcal{F}_{k,v,u}$ to $\mathcal{F}_{k-1,e,v}$. We can then write

$$\begin{aligned} h_\alpha(u, v) &= \frac{\sum_{k=1}^n \beta^k \sum_{F' \in \mathcal{F}_{k-1,e,v}} w^*(F')/w_e}{\sum_{k=1}^n \beta^k \sum_{F \in \mathcal{F}_k} w^*(F)} \\ &= \frac{\beta \sum_{k=1}^{n-1} \beta^k \sum_{F \in \mathcal{F}_{k,e,v}} w^*(F)}{w_e \sum_{k=1}^n \beta^k \sum_{F \in \mathcal{F}_k} w^*(F)}. \end{aligned}$$

The proof is complete. \square

By using the above theorems, we can then derive a relatively clean formulation for the generalized effective resistance.

Theorem 7. *For an edge $e = \{u, v\}$ in G and $\alpha \in [0, 1)$, the effective resistance $R_\alpha(u, v)$ satisfies*

$$R_\alpha(u, v) = \frac{\beta \sum_{k=1}^{n-1} \beta^k \sum_{F \in \mathcal{F}_{k,e}} w^*(F)}{w_e \sum_{k=1}^n \beta^k \sum_{F \in \mathcal{F}_k} w^*(F)}$$

where $\beta = 2\alpha/(1-\alpha)$ and $\mathcal{F}_{k,e}$ denotes the family of k -rooted forests containing e as an edge with k roots.

Proof. The proof follows from the fact that $\mathcal{F}_{k,e,u}$ and $\mathcal{F}_{k,e,v}$ are disjoint and

$$\mathcal{F}_{k,e} = \mathcal{F}_{k,e,u} \cup \mathcal{F}_{k,e,v}.$$

Therefore we have

$$\begin{aligned} R_\alpha(u, v) &= h_\alpha(u, v) + h_\alpha(v, u) \\ &= \frac{\beta \sum_{k=1}^{n-1} \beta^k \sum_{F \in \mathcal{F}_{k,e}} w^*(F)}{w_e \sum_{k=1}^n \beta^k \sum_{F \in \mathcal{F}_k} w^*(F)}. \end{aligned}$$

\square

6 Using the generalized hitting time to find sparse cuts

There are several useful properties of the hitting time h_α which we will describe.

Lemma 2. *For a subset $S \subset V$ with $S \neq \emptyset$, we have*

$$\sum_{\substack{\{u,v\} \in E \\ u \in S, v \notin S}} h_\alpha(u, v) \leq \frac{2\alpha}{1-\alpha}.$$

Proof. Since we can view $pr_\alpha(\chi_v)$ as a probability distribution, we have

$$\begin{aligned} & \sum_{\substack{\{u,v\} \in E \\ u \in S, v \notin S}} h_\alpha(u, v) \\ &= \sum_{\substack{\{u,v\} \in E \\ u \in S, v \notin S}} \left(\frac{[pr_\alpha(\chi_v)](v)}{d_v} - \frac{[pr_\alpha(\chi_v)](u)}{d_u} \right) \\ &= [pr_\alpha(\chi_v)](S) - [pr_\alpha(\chi_v)P](S). \end{aligned}$$

Since the PageRank vector $pr_\alpha(\chi_v)$ is the unique solution of the following equation,

$$pr_\alpha(\chi_v) = \alpha\chi_v + (1-\alpha)pr_\alpha(\chi_v)\frac{I+P}{2},$$

by rearranging the terms above, we have

$$pr_\alpha(\chi_v) - pr_\alpha(\chi_v)P = \frac{2\alpha}{1-\alpha}\chi_v - \frac{2\alpha}{1-\alpha}pr_\alpha(\chi_v).$$

Thus,

$$[pr_\alpha(\chi_v)](S) - [pr_\alpha(\chi_v)P](S) = \frac{2\alpha}{1-\alpha}\chi_v(S) - \frac{2\alpha}{1-\alpha}[pr_\alpha(\chi_v)](S) \leq \frac{2\alpha}{1-\alpha}.$$

□

For two distinct vertices, we consider a partition of V into two parts, one of which contains u and the other contains v . In particular, we are interested in the Cheeger ratio of such partitions defined as follows:

$$\phi_{uv} = \min_{\substack{S \subset V \\ u \in S, v \notin S}} \frac{e(S, \bar{S})}{\min\{\text{vol}(S), \text{vol}(\bar{S})\}}.$$

Theorem 8. *For two distinct vertices u and v and a constant $\phi \in (0, 1)$, suppose two sets $X = \{w : h_\alpha(w, v) \leq 0\}$ and $Y = \{w : h_\alpha(w, v) < h_\alpha(u, v)\}$ satisfy the following conditions:*

- (i) $h_\alpha(u, v) > 2\alpha/\phi\text{vol}(X)$;
- (ii) $\text{vol}(X) \leq \text{vol}(G)/2$;

(iii) $\text{vol}(Y)/\text{vol}(X) \leq (1 + \phi/2)$.

Then $\phi_{uv} \leq \phi$.

Proof. The proof is by contradiction. We first observe that $X \subset Y$. Suppose $\phi_{uv} > \phi$. Then we have

$$\begin{aligned} e(X, \bar{Y}) &\geq e(X, \bar{X}) - \text{vol}(Y \setminus X) \\ &\geq \phi \text{vol}(X) - (\phi/2)\text{vol}(X) && \text{by hypothesis and (iii)} \\ &\geq (\phi/2)\text{vol}(X) \end{aligned} \tag{15}$$

By (i) and Lemma 2, we have

$$\begin{aligned} \sum_{\substack{\{x,y\} \in E \\ x \in X, y \in \bar{X}}} h_\alpha(x, y) &\geq \sum_{\substack{\{x,y\} \in E \\ x \in X, y \in \bar{Y}}} h_\alpha(x, y) \\ &\geq e(X, \bar{Y})h_\alpha(u, v) \end{aligned}$$

since $h_\alpha(u, v)$ is the smallest value among all edges (x, y) such that $x \in X$ and $y \in \bar{Y}$. By inequality (15) and (i), the above inequality implies that

$$\sum_{\substack{\{x,y\} \in E \\ x \in X, y \in \bar{X}}} h_\alpha(x, y) > \alpha.$$

This is a contradiction and the proof is complete. \square

7 Using PageRank to estimate the effective Resistance

The effective resistance $R(u, v)$ is a very useful graph invariant. For example, the recent work by Spielman and Srivastava [14] on graph sparsification relies on the effective resistance. Here we wish to illustrate that we can use the generalized effective resistance to approximate the effective resistance. Then by using the effective approximation algorithm for computing PageRank [1] and the PageRank representation in (11) for the generalized effective resistance, we can approximate the effective resistance as a result.

Recall that for $\alpha = \beta/(2 + \beta)$, we have

$$R_\alpha(u, v) = \beta(\chi_u - \chi_v)D^{-1/2}\mathcal{G}_\beta D^{-1/2}(\chi_u - \chi_v)^T.$$

Theorem 9. *For two distinct vertices u and v in G , we have*

$$|\beta R(u, v) - R_\alpha(u, v)| \leq \frac{\beta^2}{\lambda_1^2} \left(\frac{1}{d_u} + \frac{1}{d_v} \right)$$

where λ_1 is the smallest nontrivial eigenvalue of \mathcal{L} of G and $\alpha = \beta/(2 + \beta)$.

Proof. From (3) and (8), we have

$$\begin{aligned} R(u, v) &= (\chi_v - \chi_u)D^{-1/2}\mathcal{G}D^{-1/2}(\chi_v - \chi_u)^T \\ &= \sum_{i=1}^{n-1} \frac{1}{\lambda_i} \left(\frac{\phi_i(u)}{\sqrt{d_u}} - \frac{\phi_i(v)}{\sqrt{d_v}} \right)^2. \end{aligned}$$

From equation (5), we have

$$R_\alpha(u, v) = \sum_{i=0}^{n-1} \frac{\beta}{\lambda_i + \beta} \left(\frac{\phi_i(u)}{\sqrt{d_u}} - \frac{\phi_i(v)}{\sqrt{d_v}} \right)^2.$$

Combining the above two expressions, we have, for all $\beta > 0$,

$$\begin{aligned} & |\beta R(u, v) - R_\alpha(u, v)| \\ & \leq \left(\frac{\phi_0(u)}{\sqrt{d_v}} - \frac{\phi_0(v)}{\sqrt{d_v}} \right)^2 + \sum_{i=1}^{n-1} \left(\frac{\beta}{\lambda_i} - \frac{\beta}{\lambda_i + \beta} \right) \left(\frac{\phi_i(u)}{\sqrt{d_u}} - \frac{\phi_i(v)}{\sqrt{d_v}} \right)^2 \\ & = \sum_{i=1}^{n-1} \frac{\beta^2}{\lambda_i(\lambda_i + \beta)} \left(\frac{\phi_i(u)}{\sqrt{d_u}} - \frac{\phi_i(v)}{\sqrt{d_v}} \right)^2 \end{aligned}$$

since $\phi_0 = \mathbf{1}D^{1/2}/\sqrt{\text{vol}(G)}$.

For two fixed vertices u and v , the vector f_u , defined by $f_u(i) = \phi_i(u)$, for $i = 0, \dots, n-1$, is orthogonal to f_v . This implies

$$\sum_{i=1}^{n-1} \frac{\phi_i(u)\phi_i(v)}{\sqrt{d_u d_v}} = -\frac{\phi_0(u)\phi_0(v)}{\sqrt{d_u d_v}} = -\frac{1}{\text{vol}(G)}$$

Thus we have

$$\begin{aligned} |\beta R(u, v) - R_\alpha(u, v)| & \leq \frac{\beta^2}{\lambda_1(\lambda_1 + \beta)} \left(\frac{2}{\text{vol}(G)} + \sum_{i=1}^{n-1} \left(\left(\frac{\phi_i(u)}{\sqrt{d_u}} \right)^2 + \left(\frac{\phi_i(v)}{\sqrt{d_v}} \right)^2 \right) \right) \\ & \leq \frac{\beta^2}{\lambda_1^2} \left(\frac{2}{\text{vol}(G)} + \frac{1}{d_u} - \frac{1}{\sqrt{d_u \text{vol}(G)}} + \frac{1}{d_v} - \frac{1}{\sqrt{d_v \text{vol}(G)}} \right) \\ & \leq \frac{\beta^2}{\lambda_1^2} \left(\frac{1}{d_u} + \frac{1}{d_v} \right) \end{aligned}$$

since f_u and f_v are orthonormal vectors. \square

References

1. R. Andersen, F. Chung, and K. Lang, Local graph partitioning using pagerank vectors, *Proceedings of 47th Annual IEEE Symposium on Foundations of Computer Science* (2006), pp. 475–486.

2. S. Brin and L. Page, The anatomy of a large-scale hypertextual Web search engine, *Computer Networks and ISDN Systems*, **30(1-7)** (1998), pp. 107–117.
3. A. k. Chandra, P. Raghavan, W. L. Ruzzo, R. Smolensky and P. Tiwari, The electrical resistance of a graph captures its commute time and cover time, *Proceedings of the 26th ACM Symposium on Theory of Computing* (1989), pp. 574–586.
4. F. Chung, *Spectral Graph Theory*, AMS Publication, 1997.
5. F. Chung and S.-T. Yau, Discrete Green's Functions. *Journal of Combinatorial Theory, Series A*, **91(1-2)** (2000), pp. 191–214.
6. H. Haveliwala, Topic-sensitive pagerank: A context-sensitive ranking algorithm for web search, *IEEE Trans. Knowl. Data Eng.*, **15(4)**, (2003), pp. 784–796.
7. Glen Jeh and Jennifer Widom, Scaling personalized web search. *Proceedings of the 12th World Wide Web Conference (WWW)*, (2003), pp. 271–279.
8. G. Green, An Essay on the Application of Mathematical Analysis to the Theories of Electricity and Magnetism, Nottingham, 1828.
9. S. Guattery and G. L. Miller, Graph embeddings and laplacian eigenvalues, *Journal on Matrix Analysis and Applications*, **21(3)** (2000), pp. 703–723.
10. G. Kirchhoff, Über die Auflösung der Gleichungen, auf welche man bei der Untersuchung der linearen Verteilung galvanischer Ströme geführt wird, *Ann. Phys. chem.* **72** (1847), 497–508
11. G. Kirchhoff, On the motion of electricity in conductors (Über die Bewegung der Elektrizität in Leitern), *Annalen der Physik*, Vol. **102**, (1857), pp. 529.
12. C. St. J. A. Nash-Williams, Random walks and electronic currents in networks, *Proceedings of the Cambridge Philosophical Society*, **55** (1959), pp. 181–194.
13. L. Lovász, Random walks on graphs: A survey, *Combinatorics*, Paul Erdős is Eighty **2** (1993), pp. 1–46.
14. D. A. Spielman and N. Srivastava, Graph sparsification by effective resistances, *Proceedings of the 40th ACM Symposium on Theory of Computing* (2008), pp. 563–568.