Link Analysis with Kernels[1]

Takahiko Ito Masashi Shimbo Taku Kudo[2]
Yuji Matsumoto

Graduate School of Information Science
Nara Institute of Science and Technology
Ikoma, Nara 630-0192, Japan.

October 13, 2004

[1]Correspondence to: shimbo@is.naist.jp.

[2]Present address: NTT Communication Science Laboratories, Seika, Soraku-gun, Kyoto 619-0237, Japan.

Abstract

We show that a family of graph kernels provides a unified perspective on the three measures proposed for link analysis: relatedness, global importance, and relative importance. This framework establishes relative importance as an intermediate between relatedness and global importance, in which the bias between relatedness and importance is naturally controlled by a parameter characterizing individual kernels in the family. In particular, we demonstrate that (1) the Neumann kernels due to Kandola et al. clarify the relationship between the co-citation and bibliographic coupling relatedness and Kleinberg's HITS importance, and (2) incorporating the graph Laplacian into formulation overcomes some limitations of co-citation relatedness. The property of these measures is examined with a real-world network of bibliographic citations.

Keywords: link analysis, graph kernel, relative importance, HITS.

Chapter 1

Introduction

Citations have been the major source of information for analyzing the relationship between scientific papers, authors, and journals from the early days of bibliometric studies. With the recent proliferation of hypertext documents in the web, many attempts have been made to apply the methods proposed in bibliometrics to the structural analysis of the web, and also to develop new methods based on bibliometric concepts. Two notable methods, PageRank [1] and HITS [12], have emerged from the research and are used extensively to evaluate the importance, or popularity, of web pages.

The 'relative importance' measure, recently proposed by White and Smyth [18], is defined as the 'importance of a node relative to one or more root nodes.' This definition contrasts with the importance measures as given by PageRank and HITS, which, having no particular root nodes in their models, can be interpreted as 'global' importance. Relative importance is appealing because it has a direct application in ranking web pages according to individual search queries.

White and Smyth carried out an extensive comparison of their proposed methods for evaluating relative importance. However, two issues were not addressed. First, it was not made clear whether and how relative importance differs from classical bibliometric measures of 'relatedness' between documents. For example, relatedness such as given by co-citation coupling [16] also meets the criteria for relative importance given in [18, Section 4]. Second, while most of the proposed methods have a parameter controlling the bias towards global importance, it was not analyzed how the parameter value affects the property of resulting relative importance measures.

In this paper, we present a unified framework that gives account for the three link analysis measures mentioned above: relatedness, global importance and relative importance. Our approach is based on a parametric family of kernels (Neumann kernels [9]) defining an inner product of nodes in a graph. The parameterization allows us to identify the co-citation relatedness measure as an extreme of the parameter range, and the HITS importance (authoritativeness) as the other extreme. As a consequence, relative importance fits naturally as a variation located somewhere between these two extremes. The resulting measure is thus biased towards relatedness or importance depending on this parameter.

Another topic addressed in this paper is the limitations of the standard methods for evaluating relatedness between documents, namely, co-citation and bibliographic coupling [11]. These methods are not capable of computing relatedness if the documents have no common citations (in the case of bibliographic coupling) or if they are not jointly cited by any documents (in co-citation). With appropriate parameter setting, the Neumann kernel can give a non-zero value to a pair of documents even if they have no common citation or if they are not jointly cited by other documents. However, the measure induced by the Neumann kernel under this parameterization is biased toward importance at the same time, making it not suitable for evaluating relatedness. This limitation is overcome with a different kernel-based formulation, which adopts the ideas from spectral graph theory. By introducing a new parameter to these kernels, we also obtain another relative importance measure.

Chapter 2

Preliminaries

In this section, we review the link analysis measures of relatedness, global importance, and relative importance. These measures are defined on a *citation graph*, which models networked data such as bibliographic citation networks and the web. A citation graph is a simple digraph, where nodes represent a set of entities such as scientific papers and web pages, while edges represent links between the entities; e.g., citations and hyperlinks.

We will also use the following definitions. A *weighted graph* is a simple graph whose edges are labeled with non-zero reals (called *weights*). A multigraph can be represented as a weighted graph if edge weights represent the multiplicity of corresponding edges in the multigraph. To be in line with this representation, the weight of a path is defined as the product of the weights of its constituent edges; the weight of a path thus equals the multiplicity of the corresponding path in the multigraph.

2.1 Relatedness

An assumption underpinning link analysis is that an edge between a pair of nodes signifies the nodes being in some sense related. Hence the degree of relatedness can be inferred from the proximity of nodes induced by the existence of edges and their weights (or multiplicity).

Co-citation [16] and bibliographic coupling [11] are the standard methods of computing relatedness (or similarity) between documents from a citation graph. For instance, CiteSeer uses co-citation coupling to recommend related papers to users.

Co-citation coupling defines relatedness between documents as the number of other documents citing them both. Bibliographic coupling, on the other hand, defines relatedness between two documents as the number of common references cited by the two. These measures can be formally defined in terms of a citation graph and its adjacency matrix.

Definition 2.1 Let A be the adjacency matrix of a citation graph G. The symmetric matrix $A^{\mathrm{T}}A$ is called the *co-citation matrix* of G. The *co-citation graph* of G is the weighted undirected graph induced by taking $A^{\mathrm{T}}A$ as its adjacency matrix. We define the *bibliographic coupling matrix* and the *bibliographic coupling graph* of G analogously by using AA^{T} in place of $A^{\mathrm{T}}A$.

In a citation graph whose adjacency matrix is A, the number of co-citations between nodes i and j is given by the (i, j)-element of $A^{\mathrm{T}}A$. Similarly, each element of AA^{T} represents the value of bibliographic coupling. Because these matrices are symmetric, their graph counterparts are undirected. Figure 2.1 illustrates a citation graph with the induced co-citation and bibliographic coupling graphs.

2.2 Importance

Because of the difficulty in computing importance of documents from contents, citation counts have long been used as the index of document importance.

Kleinberg's HITS [12], along with PageRank, is a more sophisticated method for evaluating document importance. HITS assigns two scores to each document, called the authority and hub scores.

Definition 2.2 Given the adjacency matrix A of a citation graph G, the HITS algorithm computes the following recursion over $n = 0, \dots$ starting with $a^{(0)} = h^{(0)} = \mathbf{e}$ where $\mathbf{e}$ is a vector of all 1's.

$$a^{(n+1)} = \frac{A^{\mathrm{T}}h^{(n)}}{|A^{\mathrm{T}}h^{(n)}|}, \quad h^{(n+1)} = \frac{Aa^{(n+1)}}{|Aa^{(n+1)}|}. \tag{2.1}$$

The i-th element of the *authority vector* $\lim_{n\to\infty} a^{(n)}$ represents the *authority score* of node i. Similarly, the *hub vector* $\lim_{n\to\infty} h^{(n)}$ gives the *hub scores*.

It is well known that subject to a mild assumption, the authority and hub vectors exist and equal the principal eigenvectors of $A^{\mathrm{T}}A$ and AA^{T}, respectively.

2.3 Relative importance

White and Smyth [18] have recently proposed a new link analysis measure called 'relative importance.' This measure is defined as the 'importance of nodes in a graph relative to one or more root nodes.' They argued that to estimate relative importance, simply applying global importance algorithms to the subgraph surrounding the root nodes does not work, because the root nodes are not given any special preference during importance computation. Hence they called for methods for computing relative importance directly.

Four scenarios for relative importance computation were considered in their work. Let V and E be the sets of nodes and edges, respectively.

- Given nodes $r, t \in V$, compute the "importance" $I(t|r)$ of t with respect to the root node r.
- Rank all nodes in $T \subset V$ with respect to a root node r. We can easily get the ranking by sorting values of $I(t|r)$ for each $t \in T$.
- Given nodes $T \subset V$ to rank, and a set of root nodes $R \subset V$, rank all nodes in T with respect to R.
- Rank all nodes in V.

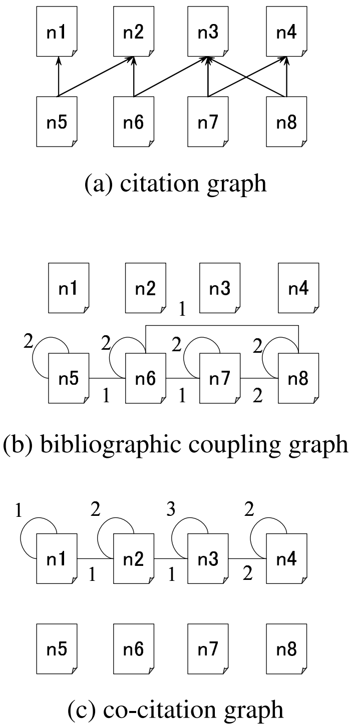

Figure 2.1: A citation graph, and the induced bibliographic coupling and co-citation graphs.

White and Smyth proposed several methods to compute relative importance under these scenarios. We will discuss some of these methods in Section 6.

Unfortunately, [18] did not make explicit what criteria must be met for a measure to qualify as relative importance, except that it should fit into one of the above scenarios and should possibly give a non-negative quantity. Consequently, the distinction is left unclear between relative importance and traditional relatedness measures such as co-citation and bibliographic coupling. For example, these appear to fit scenarios 1 through 3, if the wording "importance" is replaced with "relatedness," and if R consists only of the node against which we want to evaluate relatedness.

In this paper, we solely deal with scenarios 1 and 2, where the set R of root nodes is a singleton. This is because relative importance under multiple root nodes is rather straightforward to compute once individual importance $I(t|r)$ for each $r \in R$ is obtained. For example, White and Smyth suggested using the average of $I(t|r)$ over $r \in R$.

Chapter 3

Relative importance as a mixture of importance and relatedness

As mentioned above, one thing not addressed in the previous work on relative importance is its relationship with the traditional measure of relatedness.

Our interpretation of relative importance is that it is an intermediate between global importance of documents and their relatedness. This interpretation may seem ill-defined since global importance is a measure defined on individual nodes, whereas relatedness and relative importance are defined between them. However, given a global importance vector v such as the HITS authority and hub vectors, vv^{T} defines a matrix in which every row (and column) i such that $v(i) \neq 0$ gives a ranking of nodes identical to the one given by v. Global importance can thus be treated as a function over a pair of nodes, or a matrix, as well.

In the rest of this section and Section 4, we will develop several ways to formulate relative importance from this standpoint. These formulations are based on the family of symmetric positive semidefinite kernels, which define an inner product of nodes in a graph.

3.1 Neumann kernels

Kandola et al. [9] proposed the *Neumann kernels* for computing document similarity from terms occurring in documents, in a manner similar to latent semantic analysis. We show that these kernels can be used for computing relative importance from citation graphs as well.

The Neumann kernel in its original form is defined in terms of the document-by-term matrix X whose (i, j)-element is the frequency of the j-th term occurring in document i. As a preparation for computing the Neumann kernels, document correlation matrix $K = X^{\mathrm{T}}X$ and term correlation matrix $M = XX^{\mathrm{T}}$ are first computed. The Neumann kernel matrices are then derived from K and M as follows.

Definition 3.1 Let X be a document-by-term matrix, and let $K = X^{\mathrm{T}}X$ and $M = XX^{\mathrm{T}}$. The *Neumann kernel* matrices with *diffusion factor* $\gamma (> 0)$, denoted by $\hat{K}_\gamma$ and

$\hat{M}_\gamma$, are defined as the solution to the following system of equations.

$$\hat{K}_\gamma = \gamma X^{\mathrm{T}} \hat{M}_\gamma X + K, \quad \hat{M}_\gamma = \gamma X^{\mathrm{T}} \hat{K}_\gamma X + M. \tag{3.1}$$

The similarity between documents i and j is given by the (i, j)-element of $\hat{K}_\gamma$, and each element of $\hat{M}_\gamma$ represents the similarity between terms.

Eq. (3.1) implies an alternative representation based on the Neumann series.

$$\hat{K}_\gamma = K \sum_{n=0}^{\infty} \gamma^n K^n, \quad \hat{M}_\gamma = M \sum_{n=0}^{\infty} \gamma^n M^n. \tag{3.2}$$

Hence, when γ is less than the reciprocal of the spectral radius of K and M, the solution exists and is given by $\hat{K}_\gamma = K(I - \gamma K)^{-1}$ and $\hat{M}_\gamma = M(I - \gamma M)^{-1}$.

3.2 Relative importance based on the Neumann kernels

The recurrence over $\hat{K}$ and $\hat{M}$ in eq. (3.1) implies that the Neumann kernels evaluate similarity between documents from term similarity, and vice versa. This complementary relation is reminiscent of the recursion (2.1) between the authorities and hubs in HITS. We apply the Neumann kernels to citation analysis on the basis of this particular similarity to HITS. Specifically, we use the adjacency matrix A of a citation graph in place of document-by-term matrix X. We thus have $K = A^{\mathrm{T}} A$ and $M = AA^{\mathrm{T}}$, which coincide with the co-citation and bibliographic coupling matrices, respectively. Plugging them into eq. (3.2) yields the Neumann kernels based solely on citation information. For convenience, we introduce the following notation.

$$N_\gamma(B) = B \sum_{n=0}^{\infty} (\gamma B)^n, \tag{3.3}$$

and henceforth write

$$\hat{K}_\gamma = N_\gamma(A^{\mathrm{T}} A) = A^{\mathrm{T}} A \sum_{n=0}^{\infty} \gamma^n (A^{\mathrm{T}} A)^n, \tag{3.4}$$

$$\hat{M}_\gamma = N_\gamma(AA^{\mathrm{T}}) = AA^{\mathrm{T}} \sum_{n=0}^{\infty} \gamma^n (AA^{\mathrm{T}})^n. \tag{3.5}$$

3.3 Interpretation

We now discuss the interpretation of Neumann kernels in terms of link analysis, by using $N_\gamma(A^{\mathrm{T}} A)$ of eq. (3.4) as an example.

Eq. (3.4) gives a weighted sum of $(A^{\mathrm{T}} A)^n$ over every $n = 1, 2, \ldots$. This means that each element in the kernel matrix $N_\gamma(A^{\mathrm{T}} A)$ represents the weighted sum of the number of paths between nodes, since the term $(A^{\mathrm{T}} A)^n(i, j)$ represents the number of paths of length n between nodes i and j in the co-citation graph.

Each term $(A^{\mathrm{T}} A)^n$ can be further interpreted as follows.

When length n is small, the number $(A^{\mathrm{T}} A)^n(i, j)$ of paths between i and j is non-zero only if their distance is less than n. In other words, $(A^{\mathrm{T}} A)^n$ captures the degree of proximity between any two nodes. Since relatedness is a measure based on proximity in the graph, $(A^{\mathrm{T}}A)^n(i,j)$ can be interpreted as indicating relatedness between two nodes. In particular, $(A^{\mathrm{T}}A)^1$ is gives the co-citation matrix.

By contrast, when n is sufficiently large, the number of paths of length n between i and j indicates the importance of these nodes, as the following proposition states. See Appendix A for its proof.

Proposition 3.1 *Let λ be the principal eigenvalue of a nonnegative symmetric matrix $A^{\mathrm{T}}A$, and suppose that λ is a simple eigenvalue (i.e., λ has a multiplicity of one). There exists a unit eigenvector v corresponding to λ such that $(1/\lambda)(A^{\mathrm{T}}A)^n \to vv^{\mathrm{T}}$ as $n \to \infty$.*

The above proposition implies that when the graph induced by $A^{\mathrm{T}}A$ is connected, the ranking of nodes induced by the magnitude of the components in each row of $(A^{\mathrm{T}}A)^n$ tends towards the HITS authority ranking (given by the principal eigenvector). Formally speaking,

Corollary 3.1 *Let v be a unit eigenvector corresponding to the principal eigenvalue of $A^{\mathrm{T}}A$. For any node i with $|v(i)| > 0$ and any node pair (j,k) such that $|v(j)| > |v(k)|$, there exists a positive integer m satisfying*

$$(A^{\mathrm{T}}A)^n(i,j) > (A^{\mathrm{T}}A)^n(i,k), \qquad \forall n \geq m.$$

Putting the above discussions together, we see that summing $(A^{\mathrm{T}}A)^n$ over $n = 1, 2, \ldots$ as in eq. (3.4) can be interpreted as computing the mixture of relatedness (when n small) and importance (when n large). The Neumann kernels thus provide a measure of relative importance from the perspective stated in the beginning of this section; the (i,j)-element of $N_\gamma(A^{\mathrm{T}}A)$ or $N_\gamma(AA^{\mathrm{T}})$ gives the importance of node j relative to node i, and the bias between relatedness and global importance is determined by parameter γ. As a special case where $\gamma = 0$, the Neumann kernels subsume co-citation and bibliographic coupling.

We also have the following proposition which implies that Neumann kernels subsume the HITS global importance at the ceiling of γ. Recall that γ is subject to the constraint $\gamma < 1/\lambda$, where λ is the spectral radius of $A^{\mathrm{T}}A$.

Proposition 3.2 *Let λ be the principal eigenvalue of a nonnegative symmetric $A^{\mathrm{T}}A$, and λ is a simple. There exists a unit eigenvector v corresponding to λ such that*

$$\left(\frac{1}{\lambda} - \gamma\right) N_\gamma(A^{\mathrm{T}}A) \to vv^{\mathrm{T}} \qquad \text{as } \gamma \to \frac{1}{\lambda} - 0.$$

Proof See Appendix. □

Chapter 4

Variations with a Laplacian family of kernels

4.1 Limitations of co-citation coupling relatedness

Section 3 showed that the Neumann kernels clarify the relationship between the HITS importance, and the co-citation and bibliographic coupling relatedness. A problem with this formulation is that the induced measure inherits the limitations of co-citation and bibliographic coupling. We show in this section that these limitations can be overcome with a different kernel-based formulation. The limitations we address are as follows.

Problem 4.1 Co-citation coupling assigns a non-zero relatedness score to a pair of documents only when they are commonly referenced by a document.

In the citation graph of Figure 2.1(a), documents n_1 and n_3 are not jointly cited by any documents, and this result in an edge between n_1 and n_3 absent in Figure 2.1(c). They are hence not related to each other in terms of co-citation coupling. It is however desirable to capture a weak relationship between n_1 and n_3 as well; i.e., they are both cited by the papers (n_5 and n_6) commonly referring to n_2.

Problem 4.2 Co-citation coupling determines the relatedness between documents i and j only on the basis of the number of documents commonly citing the two. It ignores the relationship among i, j and documents not citing both i and j; e.g., the number of citations from these documents does not affect the relatedness between i and j.

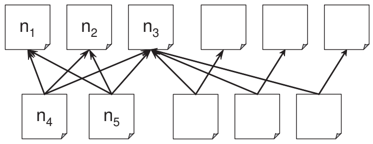

Figure 4.1: A citation graph illustrating a limitation of co-citation relatedness.

We illustrate why Problem 4.2 is an issue using the graph of Figure 4.1. In this graph, n_3 represents a frequently linked web page such as Google and Yahoo. Pages n_1 and n_2, on the other hand, do not have any other referring pages besides pages n_4 and n_5. Our intuition dictates that n_2 is more related to n_1 than to n_3, as one would not conclude a page is related (or similar) to Google or Yahoo just because it is jointly cited with them. However, because each of the page pairs (n_1, n_2) and (n_2, n_3) has two common links respectively, n_1 and n_3 are estimated as equally related to n_2 with co-citation coupling.

Although co-citation is used in the above exposition, it is easy to see that analogous limitations carry over to bibliographic coupling.

Neumann kernels, when viewed as a relatedness measure, also suffer from these limitations. They count the paths of any length between two nodes (see Section 3.3), and thus at first sight appear to alleviate Problem 4.1 if parameter γ is set sufficiently large. However, increased γ biases the induced measure towards global importance at the same time, making it unsuitable for evaluating relatedness. This bias also incurs Problem 4.2. When applied to the example of Figure 4.1, the Neumann kernels with non-zero γ assign a greater score to n_3 than to n_1 with respect to the root node n_2, contradicting our intuition on relatedness measure even further than co-citation coupling. For example, the sub-matrix of the Neumann kernel with $\gamma = 0.18$ for nodes n_1 through n_3 is shown below.

$$N_{0.18} = \begin{pmatrix} 1.89 & 3.05 & 3.40 \\ 3.05 & 8.33 & 15.49 \\ 3.40 & 15.49 & 46.12 \end{pmatrix}$$

It may seem that the desired result is achievable by simply normalizing the Neumann kernel matrix N to obtain its normalized version $\bar{N}$, defined by

$$\bar{N}(i,j) = N(i,j)/\sqrt{N(i,i)N(j,j)}.$$

Unfortunately, this scheme does not work as expected. The following submatrix of $\bar{N}_{0.18}$ still shows that $\bar{N}_{0.18}(2,1) < \bar{N}_{0.18}(2,3)$.

$$\bar{N}_{0.18} = \begin{pmatrix} 1.00 & 0.77 & 0.36 \\ 0.77 & 1.00 & 0.79 \\ 0.36 & 0.79 & 1.00 \end{pmatrix}.$$

We have also verified that the problem is not solved by either using the distance of nodes in the kernel feature spaces (both before and after normalizing the matrix), or by reweighting the adjacency matrix with tf.idf.

4.2 Extension with the graph Laplacian

We propose to use the *graph Laplacian* [2] to overcome the limitations of Section 4.1.

Definition 4.1 Let B be the adjacency matrix of an undirected graph G with positive weights. The *Laplacian* (matrix) of G is defined as

$$L(B) = D(B) - B, \tag{4.1}$$

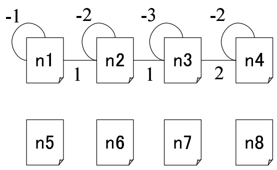

Figure 4.2: Graph induced by the negative of the Laplacian of the co-citation graph of Figure 2.1(c).

where $D(B)$ is a diagonal matrix whose diagonal elements are given by

$$D(B)(i,i) = \sum_j B(i,j). \tag{4.2}$$

Using $-L(B)$ in place of the adjacency matrix B and and dropping the first factors B from eq. (3.3) we obtain a kernel called the *regularized Laplacian kernel* [17].

Definition 4.2 Let B be a nonnegative symmetric matrix and let $\gamma > 0$. If the infinite series

$$R_\gamma(B) = \sum_{n=0}^{\infty} \gamma^n \left(-L(B)\right)^n \tag{4.3}$$

converges, it is called the *regularized Laplacian kernel* on the graph induced by B with diffusion rate γ.

To apply this kernel to link analysis, we take as B the co-citation matrix $A^{\mathrm{T}}A$ or the bibliographic coupling matrix AA^{T} for a given citation digraph A, since the matrix B must be symmetric.

In a parallel to Section 3.3, computing the regularized Laplacian kernel can be interpreted as path counting as well. However, counting takes place not in a co-citation or a bibliographic coupling graph, but in the graph induced by taking the negative of their Laplacians as the adjacency matrix. Figure 4.2 depicts an example of the graph induced by $-L(B)$, where B is the adjacency matrix of the co-citation graph of Figure 2.1(c). Comparing this figure with Figure 2.1(c) illustrates that using the Laplacian can be interpreted as modifying the weights of self-loops in graphs.

4.3 Relatedness measure induced by the regularized Laplacian kernels

Unlike the Neumann kernels, the regularized Laplacian kernels do not suffer from the limitations mentioned Section 4.1. Both kernels count all paths regardless of length, and thus assign a non-zero value to any pair of nodes as long as they are connected. Unlike the Neumann kernel, however, the regularized Laplacian kernels remain a relatedness measure even when diffusion rate γ is increased, by virtue of negative weights assigned to self-loops.

To see why, consider a pair (n_0, n_t) of nodes connected through a simple path $\pi = (n_0, n_1, \ldots, n_t)$ in the graph induced by the negative of Laplacian. Because the path is simple and all constituent edges have a positive weight, the weight of π is positive[1]. Further assume that node n_0 has a large number of co-citations and hence a self-loop with a large negative weight[2]. Without the negative-weight self-loop at n_0, π would receive a weight of γ^t in computing the inner product of n_0 and n_t. However, because the regularized Laplacian kernels counts all the path regardless of weight, they also take into account the path $\pi' = (n_0, n_0, n_1 \ldots, n_t)$ as well, which is identical to π except that n_0 is passed twice through a self-loop in the beginning. Path π' has a negative weight because the weight of the self-loop is negative and the weight of π is positive. Therefore, summing the weights of both π and π' has the net effect equivalent to discounting the weight of π. If a self-loop participating in π' has a negative weight of large magnitude, the incurred amount of discounting is also large. And a node has a self-loop with weight of large magnitude if it is an authoritative node having a large number of co-citations. The end result is that the inner product between authoritative nodes and any other nodes is significantly cut down. Note that in Neumann kernels, path π' has a positive weight and counting it actually increases the inner product of n_0 and other nodes.

Example 4.1 Consider the citation graph depicted in Figure 4.1 again. As argued in Section 4.1, it is more natural to regard n_2 as more related to n_1 than to n_3. The regularized Laplacian kernel captures this aspect, and assigns a greater relatedness score to n_1 than to n_3 relative to n_2. For instance, the regularized Laplacian kernel with $\gamma = 0.1$ in this case is given by

$$R_{0.1} = \begin{pmatrix} 0.70 & 0.15 & 0.13 \\ 0.15 & 0.70 & 0.13 \\ 0.13 & 0.13 & 0.53 \end{pmatrix}.$$

Again, only the submatrix of the kernel for nodes n_1 through n_3 is shown. Note that $R_{0.1}(2,1) > R_{0.1}(2,3)$.

4.4 Controlling bias in the regularized Laplacian kernels

While parameter γ controls the bias between importance and relatedness in Neumann kernels, it does not serve the same purpose in the regularized Laplacian kernels. We will verify this claim empirically in the experiment of Section 5.2. As an alternative way to control bias in the regularized Laplacian kernels, we introduce a new parameterization scheme.

Definition 4.3 Let B be the adjacency matrix of an undirected graph G with positive weights, and let $0 \leq \alpha \leq 1$. The *modified Laplacian* matrix $L_\alpha(B)$ of the graph G is defined as

$$L_\alpha(B) = \alpha D(B) - B,$$

where $D(B)$ is the diagonal matrix defined by eq. (4.2).

[1] Recall that the weight of a path is the *product* of the weights of its constituent edges.
[2] To simplify the discussion, we neglect self-loops in other nodes.

Definition 4.4 Let B be a nonnegative symmetric matrix, and let $\gamma \geq 0$ and $0 \leq \alpha \leq 1$. The *modified regularized Laplacian kernel* on the graph induced by B with parameters γ and α, is given by

$$R_{\gamma,\alpha}(B) = \sum_{n=0}^{\infty} \gamma^n \left(-L_\alpha(B)\right)^n,$$

where $L_\alpha(B)$ is the modified Laplacian of the graph induced by B.

When $\alpha = 1$, $R_{\gamma,\alpha}(B)$ reduces to the (original) regularized Laplacian kernel $R_\alpha(B)$, and hence their elements represent relatedness between nodes. As α decreases towards 0, each row (column) vector of the kernel matrix bears more and more the character of an importance vector, provided that γ is sufficiently large. In particular, at $\alpha = 0$, $R_{\gamma,\alpha}(B)$ reduces to $B^{-1}N_\gamma(B)$, where $N_\gamma(B)$ is the Neumann kernel with diffusion rate γ on the graph induced by B.

We have the following theorem on the positive semidefiniteness of the modified regularized Laplacian kernels, showing that they indeed define an inner product.

Proposition 4.1 *For any nonnegative symmetric matrix B, the modified regularized Laplacian kernel $R_{\gamma,\alpha}(B)$ is symmetric positive semidefinite.*

Proof See Appendix. □

4.5 Diffusion kernels

The kernels presented above are based on the Neumann series, but other series can be used as well. Using the matrix exponential in place of the Neumann series yields the so-called *diffusion (heat) kernels*, originally developed in the context of spectral graph theory [2]. It was first introduced to machine learning community by Kondor and Lafferty [13].

Definition 4.5 Let G be an undirected graph with positive weights, and B be its adjacency matrix. The *diffusion kernel* matrix H_γ on G with diffusion rate $\gamma \geq 0$ is given by

$$H_\gamma(B) = \exp(-\gamma L(B)) = \sum_{n=0}^{\infty} \frac{\gamma^n \left(-L(B)\right)^n}{n!}. \tag{4.4}$$

Because the series in eq. (4.4) is convergent regardless of γ, we do not have to worry about the convergence issue. This however does not imply all γ give a useful measure of relatedness. Too large a value of γ yields a trivial 'uniform' relatedness measure which regards all nodes in a connected component as equally related.

It is conceivable to use the modified Laplacian $L_\alpha(B) = \alpha D(B) - B$ in place of the Laplacian $L(B)$ with diffusion kernels. The resulting kernel $H_{\gamma,\alpha}(B)$ allows for controlling the bias between relatedness and importance just like the modified regularized Laplacian kernels.

Proposition 4.2 *For any nonnegative symmetric matrix B, $H_{\gamma,\alpha}(B) = \exp(-\gamma L_\alpha(B))$ is symmetric positive semidefinite.*

Proof See Appendix. $\square$

We can also show that when v is the HITS authority vector of a citation graph A, and λ is the principal eigenvector of $A^{\mathrm{T}}A$, $H_{0,\gamma}(A^{\mathrm{T}}A)\,/\,\exp(\gamma\lambda)$ converges to vv^{T} as $\gamma \to \infty$, using a technique similar to the one used for Proposition 3.2.

Chapter 5

Evaluation with bibliographic citation data

To evaluate the characteristics of the measures induced by the Neumann kernels, the regularized Laplacian kernels, and the diffusion kernels, we applied them to a network of bibliographic citations. The network consists of 2867 nodes each representing a paper in the field of natural language processing.

Using the co-citation graph built from the dataset, we computed the kernel matrices under various parameter settings. We treated kernels as ranking methods in the following way. Given a root node (paper) whose index is i, an importance ranking relative to i can be obtained from the i-th column vector in each kernel matrix; i.e., by taking the j-th component of the vector as the relative importance score of the j-th paper.

The results were examined as follows. To grasp the character of individual kernels, we compared the top-10 lists of important papers relative to a fixed root paper, namely, 'Empirical studies in discourse' by M. A. Walker and J. D. Moore, *Computational Linguistics* 23(1):1–12, 1997. For quantitative evaluation, we examined the correlations among the top-10 lists, using each paper as the root node one by one.

5.1 Neumann kernels

Table 5.1 shows the list of top-10 papers induced by the Neumann kernels for the root paper 'Empirical studies in discourse.' Columns K, C and H respectively display the rankings induced by the Neumann kernels, co-citation, and the HITS authority score. A '$-$' in column C indicates that the paper was not co-cited with the root paper, and thus it could not be ranked with co-citation coupling.

As Table 5.1(a) shows, the top-10 list for the Neumann kernel with $\gamma = 0.005$ is identical to that of HITS.

By decreasing γ to 0.0048, we see that the ranking is less similar to that of HITS. More papers co-cited with the root paper begin to slip into the list.

When γ is further dropped to 0.001 (Table 5.1(c)), the paper 'Building a large annotated corpus of English: the Penn Treebank' is finally ranked lower than the root paper. The majority is formed by the papers on discourse processing, including the root paper itself. Ranked topmost is the root paper, followed by the six papers co-cited with the root paper, as indicated by column C. In fact, no other papers are co-cited with the

Table 5.1: The top-10 list of papers induced by the Neumann kernels on the root paper 'Empirical studies in discourse'

(a) $\gamma = 0.005$

K	C	H	Title
1	2	1	Building a large annotated corpus of English: the Penn Treebank
2	–	2	A stochastic parts program and noun phrase parser for unrestricted text
3	–	3	Statistical decision-tree models for parsing
4	–	4	A new statistical parser based on bigram lexical dependencies
5	–	5	Unsupervised word sense disambiguation rivaling supervised methods
6	–	6	Word-sense disambiguation using statistical models of Roget's categories trained
7	–	7	The mathematics of statistical machine translation: parameter estimation
8	–	8	Three generative, lexicalised models for statistical parsing
9	–	9	Transformation-based error-driven learning and natural language processing: a case study in part-of-speech tagging
10	–	10	Integrating multiple knowledge sources to disambiguate word sense: an exemplar-based approach

(b) $\gamma = 0.0048$

K	C	H	Title
1	2	1	Building a large annotated corpus of English: the Penn Treebank
2	1	771	Empirical studies in discourse
3	2	50	Attention, intentions, and the structure of discourse
4	–	3	Statistical decision-tree models for parsing
5	–	4	A new statistical parser based on bigram lexical dependencies
6	–	2	A stochastic parts program and noun phrase parser for unrestricted text
7	2	76	Assessing agreement on classification tasks: the kappa statistic
8	2	201	The reliability of a dialogue structure coding scheme
9	–	8	Three generative, lexicalised models for statistical parsing
10	–	6	Word-sense disambiguation using statistical models of Roget's categories trained

(c) $\gamma = 0.001$

K	C	H	Title
1	1	771	Empirical studies in discourse
2	2	1	Building a large annotated corpus of English: the Penn Treebank
3	2	50	Attention, intentions, and the structure of discourse
4	2	76	Assessing agreement on classification tasks: the kappa statistic
5	2	201	The reliability of a dialogue structure coding scheme
6	2	604	Message Understanding Conference (MUC) tests of discourse processing
7	2	1061	Effects of variable initiative on linguistic behavior in human-computer spoken natural language dialogue
8	–	3	Statistical decision-tree models for parsing
9	–	4	A new statistical parser based on bigram lexical dependencies
10	–	96	Centering: a framework for modeling the local coherence of discourse

Table 5.2: Average K-min distance between the Neumann kernels and HITS. The row HITS (limited data) shows the result when the data are limited to the papers with a non-zero HITS score.

	Neumann kernels (γ)				
	0.005	0.0045	0.004	0.0035	0.003
HITS	20.4	81.7	86.2	87.8	88.7
HITS (limited data)	0.0	76.9	82.7	84.7	85.8

root paper in the dataset. Note further that these six papers are ranked in the order of the HITS score.

To sum up, this result exemplify the role of γ we stated in Section 3.3; as parameter γ decreases towards 0, the measures induced by the Neumann kernels deviate from global importance towards relatedness.

Table 5.2 shows the result of quantitative evaluation carried out to confirm this trend. It shows the minimizing Kendall (K-min) distances [6] between the top-10 lists induced by the Neumann kernels and HITS, averaged over all root nodes. A small K-min distance means the two rankings are similar; it is equal to 0 if all top-10 items are identical, and takes the maximum value of 100 if there are no common items in the top-10 lists.

The ranking of Neumann kernels is indeed biased towards the HITS ranking as γ is increased. In particular, excluding the papers having a zero HITS authoritative score from the dataset (see the row marked 'HITS (limited data)' in the table), we see that the rankings given by the Neumann kernel at $\gamma = 0.005$ and HITS are identical.

5.2 Regularized Laplacian and diffusion kernels

Table 5.3 lists the top-10 papers induced by the regularized Laplacian kernel with $\gamma = 0.001$ relative to 'Empirical studies in discourse.'

Although there are only seven papers (including the root paper) having a non-zero co-citation relatedness score, the output of the regularized Laplacian kernel is not limited to seven. All the extra three papers ranked 8th to 10th deal with the same topic as the root paper, namely, discourse processing, as the titles of these papers show. This result is contrastive to that of the Neumann kernel discussed in Section 5.1. The Neumann kernel with $\gamma = 0.001$ (Table 5.1(c)) also lists more than seven papers, but two of the three additional papers deal not with discourse processing but with parsing. These two papers are also high-ranked papers according to HITS, and hence exemplify the property of the Neumann kernels that their ranking is biased towards global importance even when diffusion rate γ is relatively small.

By contrast, the regularized Laplacian kernels is less prone to the HITS ranking. This can also be seen from Table 5.5 listing the average K-min distances among the rankings of the regularized Laplacian kernels and HITS. The rankings of these kernels

Table 5.3: Output of the regularized Laplacian kernel with $\gamma = 0.001$ on 'Empirical studies in discourse'.

K	C	H	Title
1	1	771	Empirical studies in discourse
2	2	1061	Effects of variable initiative on linguistic behavior in human-computer spoken natural language dialogue
3	2	604	Message Understanding Conference (MUC) tests of discourse processing
4	2	201	The reliability of a dialogue structure coding scheme
5	2	76	Assessing agreement on classification tasks: the kappa statistic
6	2	50	Attention, intentions, and the structure of discourse
7	2	1	Building a large annotated corpus of English: the Penn Treebank
8	–	198	A prosodic analysis of discourse segments in direction-giving monologues
9	–	96	Centering: a framework for modeling the local coherence of discourse
10	–	115	Combining multiple knowledge sources for discourse segmentation

Table 5.4: Output of the diffusion kernel with $\gamma = 0.1$ on 'Empirical studies in discourse'.

K	C	H	Title
1	1	771	Empirical studies in discourse
2	2	1061	Effects of variable initiative on linguistic behavior in human-computer spoken natural language dialogue
3	2	604	Message Understanding Conference (MUC) tests of discourse processing
4	2	201	The reliability of a dialogue structure coding scheme
5	2	76	Assessing agreement on classification tasks: the kappa statistic
6	–	1054	Preventing false inferences
7	–	1046	Experiments in evaluating interactive spoken language systems
8	–	685	Discourse segmentation by human and automated means
9	–	780	A proposal for incremental dialogue evaluation
10	–	1026	Human-machine problem solving using spoken language systems (SLS): factors affecting performance and user satisfaction

Table 5.5: Average K-min distance among the regularized Laplacian kernels, diffusion kernels, and HITS.

		Regularized Laplacian			Diffusion		
		$\gamma = 0.001$	0.0005	0.0001	$\gamma = 0.1$	0.01	0.001
Regularized Laplacian	$\gamma = 0.001$	0.0	3.6	3.9	26.9	8.2	3.7
	0.0005		0.0	1.9	27.0	8.6	3.4
	0.0001			0.0	27.1	8.7	3.4
Diffusion	$\gamma = 0.1$				0.0	33.6	36.6
	0.01					0.0	21.8
	0.001						0.0
HITS		95.9	95.9	96.5	99.9	99.4	96.6

do not resemble that of HITS even when γ is increased to 0.001, a value nearly equal the ceiling of the admissible range for eq. (3.1) to have a solution. The table also shows that the variation in the value of γ gives only a slight modification to the rankings of the regularized Laplacian kernels.

The correlation between the regularized Laplacian kernels and co-citation coupling was also examined. For every root paper in the dataset, the set of k co-cited papers matches those ranked in the top-k list induced by the regularized Laplacian kernels with $\gamma = 0.0001$ and 0.001.

These results, when put together, support the claim that the regularized Laplacian kernel is a relatedness measure regardless of the value of γ.

The same trend applies to the diffusion kernels. The ranking of the diffusion kernels with $\gamma = 0.001$ for 'Empirical studies in discourse' is identical to that of the regularized Laplacian kernel with the same γ, shown in Table 5.3. By increasing γ to 0.1 (Table 5.4), we see that the diffusion kernels more severely undervalue authoritative papers, such as 'Building a large annotated corpus of English: the Penn Treebank.' Although this paper is co-cited with the root paper, it is left out of the top-10 list in this case. Meanwhile, all papers that crept into the list in place are concerned with discourse processing.

5.3 The modified regularized Laplacian kernels

We now examine the effect of parameter α on the modified regularized Laplacian kernels. To see if this parameter controls the tradeoff between relatedness and global importance, we compare the rankings induced by these kernels with those of HITS and the regularized Laplacian kernel with $\gamma = 0.001$. The latter two are used as the benchmark of the global importance and relatedness measures, respectively. We again use the average K-Min distance to assess the correlation. In this experiment, the diffusion rate γ is set nearly equal to the maximum admissible value $\| - L_\alpha \|_2^{-1}$ for each α, where L_α is the modified Laplacian kernel on the co-citation graph.

The result is shown in Table 5.6. The modified regularized Laplacian kernel tends to be more similar to the HITS ranking as α is decreased. Confining the dataset only to the papers having a non-zero HITS authority score (see the row marked 'HITS (limited data)'), we see that the top-10 ranking induced by the modified regularized Laplacian kernel with $\alpha = 0.01$ is identical to that of HITS.

On the other hand, when α is large, the rankings of the modified regularized Laplacian kernels are similar to relatedness measure induced by the unmodified regularized Laplacian kernel.

Table 5.6: Average K-min distance between the modified regularized Laplacian kernels and other link analysis measures: HITS, and the unmodified regularized Laplacian kernel with $\gamma = 0.001$.

	Modified regularized Laplacian (α)			
	0.01	0.0316	0.1	0.316
HITS	20.4	27.6	48.3	95.8
HITS (limited data)	0.0	9.0	35.0	94.7
Regularized Laplacian	92.6	92.6	92.8	29.5

Chapter 6

Discussion and related work

6.1 Unified framework for link analysis measures

Most of the previous work on link analysis has treated relatedness and importance as independent measures. One exception is the work by Ding et al. [4] who pointed out the relation between HITS and co-citation and bibliographic coupling. However, because their objective was unifying importance ranking algorithms such as PageRank and HITS, the existence of an interpolated measure between relatedness and importance was not addressed.

6.2 Relationship between link analysis and kernels

Smola and Kondor [17] pointed out the connection between kernels on graphs and link analysis methods including PageRank and HITS. However, only the connection to the global importance measures was addressed, and relatedness and relative importance measures were left out.

Moreover, their formulation appears quite different from ours. Smola and Kondor state that for a given node in a regular graph, its HITS score is given by the length of a vector corresponding to the node in a kernel-induced feature space of a Laplacian-based kernel. Our formulation, by contrast, obtains global importance scores using adjacency matrices instead of the Laplacian. The idea of modifying the diagonal elements of the Laplacian matrix is not present in Smola and Kondor's formulation either. Clarifying the relationship between their formulation and ours is an issue we plan to address in the future.

Saerens and his colleagues [7, 14] have recently proposed to use the *average commute time* as a distance measure between nodes in a graph. This quantity is based on the Markov-chain model of random walk, and is shown to be efficiently computable from the pseudoinverse of the Laplacian. They also prove this pseudoinverse is a positive semidefinite kernel.

6.3 Computational complexity

A major drawback of our approach is that it is computationally intensive; as it requires computing the entire kernel matrix and involves matrix inversion or exponentiation to do so, their computational complexity is O($|V|^3$) and the storage space requirement is O($|V|^2$), where $|V|$ is the number of nodes in the graph.

We plan to investigate the approximated way to compute elements of the kernel matrix on demand, by explicitly enumerating the paths between a given pair of nodes. This is described in the next subsection.

6.4 Path-based computation of relative importance

A method proposed by White and Smyth [18], which they call K short path, defines a relative importance measure as the sum of the weight of paths between the two nodes.

This idea is compatible with our view of the Neumann kernels as path counting; kernels essentially compute the sum of the weight of the paths between two nodes.

It may be possible to develop a path-based method which approximates the kernels to save time and space on the basis of this idea. Many efficient algorithms, e.g., [5], exist for enumerating paths between two nodes in order of path length or cost.

6.5 k-step Markov approach and Katz similarity

Of the algorithms for relative importance proposed by White and Smyth, the most similar to our formulation is the *k-step Markov approach*.

Given the Markov transition matrix $\tilde{A}$ induced from a citation graph $G = (V, E)$, let $\tilde{v}_0$ be a vector of initial probabilities for the root set $R \subset V$. The relative importance vector $\tilde{v}$ given by the k-step Markov approach is defined as

$$\tilde{v} = \tilde{A}\tilde{v}_0 + \tilde{A}^2\tilde{v}_0 + \cdots + \tilde{A}^k\tilde{v}_0,$$

where the i-th element of $\tilde{v}$ represents the importance of node i relative to the root node set R.

Although this method is similar to Neumann kernels as the above equation shows, the k-step Markov approach only uses the paths of length at most k, and the transition matrix of the original citation graph, which is generally asymmetric. Neumann kernels, by contrast, operates on the co-citation and bibliographic coupling matrices and sum the path weights over all paths regardless of length.

The Katz similarity[1] [10] is another method similar to the Neumann kernels and k-step Markov approach. It is defined as the infinite series

$$Z_\gamma(A) = \sum_{n=1}^{\infty} (\gamma A)^n = (I - \gamma A)^{-1} - I,$$

where A is an adjacency matrix and γ is a diffusion rate parameter. Since $Z_\gamma(A) = \gamma N_\gamma(A)$, the Katz similarity $Z_\gamma(A)$ is also a symmetric positive semidefinite kernel provided that A is symmetric.

6.6 Combining with different kernels

Since our proposed measures are defined as kernels, they can be easily combined with other kernels to obtain a new one [3, 15]. For example, Joachims at al. [8] improve the

[1] We thank Marco Saerens for pointing out to us the resemblance between the Neumann kernels and the Katz similarity.

performance of document classification by combining two kernels each incorporating different prior knowledge.

Chapter 7

Conclusions

We have argued that kernel methods provide a unified perspective on three measures for link analysis: relatedness, global importance, and relative importance. In this perspective, relative importance is interpreted as a mixture of importance and relatedness.

Several types of kernel-based formulation and their characteristics were examined.

The Neumann kernels, when applied to citation graphs, yield a measure of relative importance intermediate between co-citation/bibliographic coupling and the HITS importance measures, providing a new account of the relationship between these well-known measures. The parameter of the Neumann kernels allows us to tune the bias of the induced measure between relatedness and importance.

The limitations of the co-citation and bibliographic coupling relatedness can be overcome with the kernels based on the Laplacian of graphs, including the regularized Laplacian and diffusion kernels. Although the measures induced by these kernels are useful for evaluating node relatedness, it turned out that this formulation does not permit us to obtain either global or relative importance measures in a manner similar to the Neumann kernels. As a remedy, we have proposed an alternative parameterization for controlling the tradeoff between relatedness and importance, i.e., by modifying the diagonal elements of Laplacian.

The characteristics of these kernels were examined with real bibliographic citation data.

Appendix A

Proofs

Lemma A.1 *Let $A \in \mathbb{R}^{m\times m}$ be a non-negative square matrix, and let $\{\lambda_i \mid i = 1,\dots,m\}$ be the multiset of eigenvectors of $A^{\mathrm{T}}A$. There exists an orthonormal set of eigenvectors $\{v_i \in \mathbb{R}^m \mid i = 1,\dots,m\}$ such that*

$$\left(A^{\mathrm{T}}A\right)^n = \sum_{i=1}^{m} \lambda_i^n v_i v_i^{\mathrm{T}}, \qquad n = 1,2,\dots$$

Proof Due to the symmetry of $A^{\mathrm{T}}A$, there exists an orthonormal set of eigenvectors $\{v_i \mid i = 1,\dots,m\}$ such that for each i, v_i corresponds to eigenvalue λ_i. A spectral decomposition of $A^{\mathrm{T}}A$ is then given by

$$A^{\mathrm{T}}A = \sum_{i=1}^{m} \lambda_i v_i v_i^{\mathrm{T}}.$$

By noting that $v_i^{\mathrm{T}}v_j = 1$ if $i = j$ and is otherwise 0, we have

$$\left(A^{\mathrm{T}}A\right)^n = \sum_{i=1}^{m} \lambda_i^n v_i v_i^{\mathrm{T}}, \qquad n = 1,2,\dots. \qquad \square$$

Proposition A.1 (Proposition 3.1) *Let λ be the principal eigenvalue of a nonnegative symmetric matrix $A^{\mathrm{T}}A$, and suppose that λ is a simple eigenvalue (i.e., λ has a multiplicity of one). There exists a unit eigenvector v corresponding to λ such that $(1/\lambda)(A^{\mathrm{T}}A)^n \to vv^{\mathrm{T}}$ as $n \to \infty$.*

Proof Since $A^{\mathrm{T}}A$ is positive semidefinite, all its eigenvalues are nonnegative. Let $\{(\lambda_i, v_i) \mid i = 1,\dots,m\}$ be the set of eigenvalue-eigenvector pairs of $A^{\mathrm{T}}A$ as given in Lemma A.1, with $\lambda = \lambda_1$ the principal eigenvector and $v = v_1$. It follows that $\lambda_1 > \lambda_i \geq 0$ for all $i = 2,\dots,m$ because $\lambda = \lambda_1$ is simple by assumption. By Lemma A.1,

$$\left(A^{\mathrm{T}}A\right)^n = \lambda_1^n v_1 v_1^{\mathrm{T}} + \sum_{i=2}^{m} \lambda_i^n v_i v_i^{\mathrm{T}}, \qquad n = 1,2,\dots.$$

Dividing both sides by λ_1^n yields

$$\left(\frac{A^{\mathrm{T}}A}{\lambda_1}\right)^n = v_1 v_1^{\mathrm{T}} + \sum_{i=2}^{m} \left(\frac{\lambda_i}{\lambda_1}\right)^n v_i v_i^{\mathrm{T}}.$$

Because $\lambda_1 > \lambda_i$ for every $i = 2, \ldots, m$, the second term on the rhs vanishes as $n \to \infty$. □

Proposition A.2 (Proposition 3.2) *Let λ be the principal eigenvalue of a nonnegative symmetric $A^{\mathrm{T}}A$, and λ is a simple. There exists a unit eigenvector v corresponding to λ such that*

$$\left(\frac{1}{\lambda} - \gamma\right) N_\gamma(A^{\mathrm{T}}A) \to vv^{\mathrm{T}} \qquad \textit{as } \gamma \to \frac{1}{\lambda} - 0.$$

Proof Let $\{(\lambda_i, v_i) \mid i = 1, \ldots, m\}$ be the set of eigenvalue-eigenvector pairs as given in Lemma A.1, with $\lambda = \lambda_1$ the principal eigenvector and $v = v_1$. By eq. (3.4) and Lemma A.1,

$$N_\gamma(A^{\mathrm{T}}A) = \sum_{i=1}^{m} \left(\sum_{n=1}^{\infty} \gamma^{n-1} \lambda_i^n \right) v_i v_i^{\mathrm{T}} = \sum_{i=1}^{m} \frac{\lambda_i}{1 - \gamma\lambda_i} v_i v_i^{\mathrm{T}}.$$

After substitution,

$$\left(\frac{1}{\lambda_1} - \gamma\right) N_\gamma(A^{\mathrm{T}}A) = v_1 v_1^{\mathrm{T}} + \sum_{i=2}^{m} \frac{\lambda_i(1 - \gamma\lambda_1)}{\lambda_1(1 - \gamma\lambda_i)} v_i v_i^{\mathrm{T}}.$$

It follows that

$$\left(\frac{1}{\lambda_1} - \gamma\right) N_\gamma(A^{\mathrm{T}}A) \to v_1 v_1^{\mathrm{T}} \quad \text{as } \gamma \to \frac{1}{\lambda_1} - 0. \quad \square$$

Lemma A.2 *For any function f such that $f(A)$ is convergent, $f(A)$ is symmetric positive semidefinite if A is symmetric and $f(\lambda)$ is nonnegative for all eigenvalues λ of A.*

Proof Symmetric matrix A can be written as $A = P^{\mathrm{T}}DP$, where P is an orthogonal matrix and D is a diagonal matrix with all its diagonals being eigenvalues of A. Hence $f(D(i,i)) \geq 0$ for all $i = 1, \ldots, m$. We also have $f(A) = P^{\mathrm{T}} f(D) P$, where $f(D)$ is a diagonal matrix with $f(D)(i,i) = f(D(i,i))$. Therefore, $f(D)$ is a nonnegative diagonal matrix, and $C = f(D)^{1/2}$ exists. Substituting $f(D) = C^{\mathrm{T}}C$ in $f(A)$, we have $f(A) = P^{\mathrm{T}}C^{\mathrm{T}}CP = (CP)^{\mathrm{T}}(CP)$. □

Proposition A.3 (Proposition 4.1) *For any nonnegative symmetric matrix B, the modified regularized Laplacian kernel $R_{\gamma,\alpha}(B)$ is symmetric positive semidefinite.*

Proof Let $f(x) = \sum_{n=0}^{\infty}(-\gamma x)^n$. The existence of $R_{\gamma,\alpha}(B)$ implies $|\gamma\lambda| < 1$ for any eigenvalue λ of $L_\alpha(B)$. Hence we have $f(\lambda) = (1 + \gamma\lambda)^{-1} \geq 0$ for any eigenvalue λ of $L_\alpha(B)$. Now that $f(L_\alpha(B)) = R_{\gamma,\alpha}(B)$ by Definition 4.4 and $L_\alpha(B)$ is symmetric, the symmetric positive semidefiniteness of $R_{\gamma,\alpha}(B)$ follows from Lemma A.2. □

Proposition A.4 (Proposition 4.2) *For any nonnegative symmetric matrix B, $H_{\gamma,\alpha}(B) = \exp(-\gamma L_\alpha(B))$ is symmetric positive semidefinite.*

Proof First observe that $\exp(A)$ is positive semidefinite for any symmetric matrix A, which follows from Lemma A.2 by setting $f(A) = \exp(A)$ and the fact that $f(x) = \exp(x) \geq 0$ for any $x \in \mathbb{R}$. Since $-\gamma L_{\gamma,\alpha}(B)$ is symmetric, $H_{\gamma,\alpha}(B) = f(-\gamma L_{\gamma,\alpha}(B))$ is symmetric positive semidefinite. □

Acknowledgments

We thank Marco Saerens for interesting discussions on link analysis.

Bibliography

[1] S. Brin and L. Page. The anatomy of a large-scale hypertextual (web) search engine. *Computer Network and ISDN Systems*, 30(1–7):107–117, 1998.

[2] F. R. K. Chung. *Spectral Graph Theory*. CBMS Regional Conference Series in Mathematics 92. American Mathematical Society, Providence, RI, USA, 1997.

[3] N. Cristianini, J. Kandola, A. Elisseeff, and J. Shawe-Taylor. On optimizing kernel alignment. Technical Report NC-TR-01-087, NeuroCOLT, 2001.

[4] C. Ding, X. He, P. Husbands, H. Zha, and H. Simon. PageRank, HITS and a unified framework for link analysis. Technical Report 49372, Lawrence Berkeley National Laboratory, 2001.

[5] D. Eppstein. Finding the k shortest paths. *SIAM Journal on Computing*, 28(2):652–673, 1998.

[6] R. Fagin, R. Kumar, and D. Sivakumar. Comparing k lists. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms*, pages 28–36, 2003.

[7] F. Fouss, A. Pirotte, J.-M. Renders, and M. Saerens. A novel way of computing dissimilarities between nodes of a graph, with application to collaborative filtering. In *ECML/PKDD Workshop on Statistical Approaches to Web Mining*, pages 26–37, 2004.

[8] T. Joachims, N. Cristianini, and J. Shawe-Taylor. Composite kernels for hypertext categorization. In *Proceedings of the 18th International Conference on Machine Learning*, 2001.

[9] J. Kandola, J. Shawe-Taylor, and N. Cristianini. Learning semantic similarity. In *Advances in Neural Information Processing Systems 15*, pages 673–680. MIT Press, 2003.

[10] L. Katz. A new status index derived from sociometric analysis. *Psychmetrika*, 18(1):39–43, 1953.

[11] M. M. Kessler. Bibliographic coupling between scientific papers. *American Documentation*, 14(1):10–25, 1963.

[12] J. M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, pages 604–632, 1999.

[13] R. Kondor and J. Lafferty. Diffusion kernels on graphs and other discrete input spaces. In *Proceedings of the 18th International Conference on Machine Learning*, pages 21–24, 2001.

[14] M. Saerens, F. Fouss, L. Yen, and P. Dupont. The principal component analysis of a graph, and its relationship to spectral clustering. In *Proceedings of the 15th European Conference on Machine Learning*, pages 371–383. Springer, 2004.

[15] B. Schölkopf and A. J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge, MA, USA, 2002.

[16] H. Small. Co-citation in the scientific literature: a new measure of the relationship between two documents. *Journal of American Society for Information Science*, 24:265–269, 1973.

[17] A. J. Smola and R. Kondor. Kernels and regularization of graphs. In *Learning Theory and Kernel Machines: 16th Annual Conference on Learning Theory and 7th Kernel Workshop, Proceedings*, Lecture Notes in Artificial Intelligence 2777, pages 144–158. Springer, 2003.

[18] S. White and P. Smyth. Algorithms for estimating relative importance in networks. In *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 266–275, 2003.