Exploring Multiple Communities with Kernel-Based Link Analysis

Takahiko Ito[1], Masashi Shimbo[1], Daichi Mochihashi[2], and Yuji Matsumoto[1]

[1] Graduate School of Information Science Nara Institute of Science and Technology
{takahi-i, shimbo, matsu}@is.naist.jp
[2] ATR Spoken Language Communication Research Laboratories
daichi.mochihashi@atr.jp

Abstract. We discuss issues raised by applying von Neumann kernels to graphs with multiple communities. Depending on the parameter setting, Kandola et al.'s von Neumann kernels can identify not only nodes related to a given node but also the most important nodes in a graph. However, when von Neumann kernels are biased towards importance, top-ranked nodes are the important nodes in the dominant community of the graph irrespective of the communities where the target node belongs. To solve this "topic-drift" problem, we apply von Neumann kernels to the weighted graphs (community graph), which are derived from a generative model of links.

1 Introduction

Link analysis techniques are useful for mining knowledge from graph-structured data such as WWW and citation networks. Researchers have been trying to establish measures for evaluating the importance of individual nodes (documents) in a graph, and PageRank and HITS [1] are the two most popular 'global importance' measures. A different type of link analysis measures, the 'relatedness' between graph nodes, has been studied in bibliometrics. Co-citation coupling [2] is a classical measure of relatedness still widely used.

Our previous work [3] showed that it is possible to define a 'mixture' between the HITS global importance and co-citation relatedness through the parameterization of (von) Neumann kernels [4]. These kernels enable to compute the importance of nodes relative to individual 'root' nodes, where the degree of relativity is controlled by the parameter of the kernels.

The Neumann kernels enjoy the properties of HITS, while at the same time they inherit the problem of HITS called 'topic drifts' [5]. This problem is noticeable when the graph consists of multiple communities each addressing different topics. If the Neumann kernels are biased towards importance and applied to a multi-community graph, they assign the highest scores to the nodes in the dominant community irrespective of the root node.

This paper proposes a method to avoid topic drifts when we apply the Neumann kernels to multi-community graphs. To this end, we model the generative process of links, and construct distinct graphs for individual communities. Edges in these graphs have the weights determined by the generation probability of the citation in the respective

J. Fürnkranz, T. Scheffer, and M. Spiliopoulou (Eds.): PKDD 2006, LNAI 4213, pp. 235–246, 2006.
© Springer-Verlag Berlin Heidelberg 2006

community. Applying Neumann kernels to the community graphs, we can take communities into consideration even when we bias Neumann kernels to importance.

We also discuss the connection between our proposed kernels and the related methods, including pHITS [6] and Hofmann's Fisher kernels based on pLSI [7].

2 Preliminaries

This section reviews the link analysis measures relevant to the subsequent discussion.

Co-citation Coupling Relatedness. Co-citation [2] is the standard methods of computing relatedness (or similarity) between nodes in a citation graph. Co-citation coupling defines relatedness between documents as the number of other documents citing them both. Given the adjacency matrix A of a citation graph, the number of co-citations between nodes i and j is given by the (i, j)-element of the *co-citation matrix* $A^{\mathrm{T}}A$.

HITS Importance. Kleinberg's HITS [1], along with PageRank, is probably one of the most popular methods for evaluating document importance. HITS assigns two scores to each document, called the authority and hub scores. Let A be the adjacency matrix of a citation graph. The HITS algorithm computes the following recursion over $n = 0, 1, \ldots$ starting from $\mathbf{a}_{(0)} = \mathbf{h}_{(0)} = \mathbf{1}$.

$$\mathbf{a}_{(n+1)} = A^{\mathrm{T}}\mathbf{h}_{(n)}/|A^{\mathrm{T}}\mathbf{h}_{(n)}|, \quad \mathbf{h}_{(n+1)} = A\mathbf{a}_{(n+1)}/|A\mathbf{a}_{(n+1)}|.$$

The i-th component of the *authority vector* $\lim_{n\to\infty} \mathbf{a}_{(n)}$ represents the *authority score* of node i. Similarly, the *hub vector* $\lim_{n\to\infty} \mathbf{h}_{(n)}$ gives the *hub scores*. It is well known that under a mild assumption, the authority and hub vectors exist and equal the principal eigenvectors of $A^{\mathrm{T}}A$ and AA^{T}, respectively.

Neumann Kernels. In our previous work [3][8], we analyzed the properties of Kandola et al.'s Neumann kernels [4] as a link analysis measure.

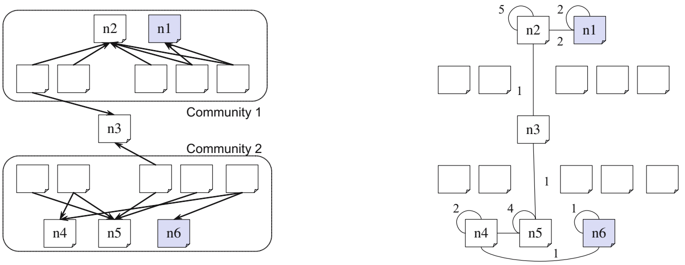
The *Neumann kernel* matrices $\hat{K}_\gamma$ and $\hat{M}_\gamma$, with a parameter γ ($0 \le \gamma < 1$) called *diffusion factor*, are defined by the following equations.

$$\hat{K}_\gamma = \sum_{n=1}^{\infty} (\frac{\gamma}{\lambda})^{n-1} \left(A^{\mathrm{T}}A\right)^n, \quad \hat{M}_\gamma = \sum_{n=1}^{\infty} (\frac{\gamma}{\lambda})^{n-1} \left(AA^{\mathrm{T}}\right)^n \tag{1}$$

Here, λ represents the dominant eigenvalue of a nonnegative symmetric matrix $A^{\mathrm{T}}A$.

Eq. (1) shows that the Neumann kernel matrix $\hat{K}_\gamma$ (or $\hat{M}_\gamma$) is a weighted sum of $(A^{\mathrm{T}}A)^n$ (or $(AA^{\mathrm{T}})^n$) over $n = 1, 2, \ldots$. Given that the (i, j)-element of the term $(A^{\mathrm{T}}A)^n$ represents the number of paths of length n between nodes i and j in the co-citation graph, we see that each element of the kernel matrix equals the weighted sum of the number of paths between nodes.

We showed in [3] that Eq. (1) can be interpreted as the mixture of co-citation relatedness and HITS importance. As a special case, the Neumann kernels $\hat{K}_\gamma$ subsume co-citation at $\gamma = 0$. At the ceiling of $\gamma \simeq 1$, on the other hand, the rankings induced by any rows of the $\hat{K}_\gamma$ are identical to the HITS importance ranking.



(a) Citation graph with multi-communities (b) Co-citation graph derived from Fig 1 (a)

Fig. 1. A citation graph with multi-communities, and the induced co-citation graphs

3 Topic Drift Problem and Neumann Kernels

HITS is known to suffer from the problem called ‘topic drifts’ [5]. If applied to a graph with multiple communities[1], HITS assigns the highest scores to the documents in the dominant community of the graph. It follows that the highest scores are assigned to documents unrelated to user’s interest, if the topic of dominant community is unrelated to the queries from users.

Consider the graph of Fig 1 (a), which contains two communities. The HITS authority ranking of this graph is $n_2 > n_1 > n_3 > n_5 > n_4 > n_6$, and we see that ranks of documents in Community 2, namely n_4, n_5 and n_6, are uniformly lower than those of documents in Community 1 (n_1 and n_2).

The Neumann kernels reduces to HITS when $\gamma \simeq 1$. This also means that they inherit the issue of topic drifts from HITS. Fig 1 (b) depicts the co-citation graph induced by the graph of Fig 1 (a). The Neumann kernels matrix $\hat{K}_{0.99}$ is shown below.

$$\hat{K}_{0.99} = \begin{pmatrix} 108.53 & 225.98 & 59.64 & 7.16 & 29.30 & 1.36 \\ 225.98 & 477.37 & 127.64 & 15.33 & 62.70 & 2.90 \\ 59.64 & 127.64 & 37.87 & 5.30 & 21.67 & 1.00 \\ 7.16 & 15.33 & 5.30 & 5.16 & 7.34 & 2.17 \\ 29.30 & 62.70 & 21.67 & 7.34 & 23.74 & 1.39 \\ 1.36 & 2.90 & 1.00 & 2.17 & 1.39 & 1.60 \end{pmatrix} . \tag{2}$$

Only the sub-matrix of the kernel matrix for documents n_1 through n_6 is shown. The remaining rows and columns are omitted because these rows and columns correspond to the isolated nodes in Fig 1 (b) and hence all their elements are 0 in the matrix.

The (i, j)-element of this matrix represents the importance of j-th document relative to the i-th document. For example, relative to document n_3 (third row), document n_2 is the most important document because the (3,2)-element (127.64) is the largest in the third row of the matrix.

[1] To the best of our knowledge, there seems to be no consensus on what constitutes ‘community’ in the literature. Roughly speaking, a ‘community’ in this paper is a set of documents in a graph citing more documents within the community than those outside.

Now let us focus on the 6-th row in the matrix. In this row, n_2 in Community 1 has the largest value. However, this ranking for n_6 is different from our intuition. The importance score of n_5, a node in the same community as n_6, should be higher than n_2, because the community in n_2 is different from that of n_6.

If we increase a diffusion factor (e.g., $\gamma = 0.999$), the ranking relative to each document will eventually be identical to that of the HITS authority ranking and determined irrespective of the community where the document belongs.

To prevent the importance rankings from diverting away from the topic (or community) users are interested in, Kleinberg [1] proposed to first extract documents that include query terms, and apply the HITS algorithm to the subgraph of the extracted documents.

We may also apply the Neumann kernels to the subgraph of documents containing query terms. However, depending on the type of data (e.g., citation networks), the document contents are not always available. Moreover, Bharat et al. [5] pointed out that there can be a discrepancy between queries and the topics of high-ranked documents, even if HITS is applied to the subgraph of documents containing query terms.

4 Proposed Method

To alleviate topic drifts in the Neumann kernels, we model the generative process of citations, and apply the Neumann kernels to the weighted graphs induced by the model. Unlike Kleinberg's solution, document contents are not used to derive these graphs.

4.1 Generative Model for Citations

Probabilistic Latent Semantic Indexing [9] (pLSI) is a method of modeling the generative process of documents. In pLSI, the joint probability between document d_i and word w_j is given by the following equation.

$$p(d_i, w_j) = \sum_{t=1}^{N} p(t)p(d_i|t)p(w_j|t) \tag{3}$$

where $t \in \{1, 2, \ldots, N\}$ represents a hidden *topic* of documents. The maximum likelihood parameters for $p(t)$, $p(d_i|t)$ and $p(w_j|t)$ are estimated by the EM algorithm.

Cohn et al. [6] modeled the generative process of citations in a similar manner to pLSI. Their model computes the generation probability of citations by using citations in place of words in Eq.(3). Thus, the probability that document i cites j is

$$p(d_i, c_j) = \sum_{t=1}^{N} p(t)p(d_i|t)p(c_j|t)$$

where d_i represents a citation emanating from document i, and c_j represents a citation to document j.

They also proposed to use $p(c_j|t)$, which is the generative probability of a citation within Community t, as the importance of document j in Community t. Although this method (pHITS) gives the importance rankings by taking the communities into account, it cannot compute the relatedness or relative importance among documents such as those given by the Neumann kernels.

4.2 Applying Neumann Kernels to Community Graphs

In this section, we propose a method to solve the topic drift problem in the Neumann kernels. The proposed method maintains the property of the Neumann kernel as a mixture of importance and relatedness, and also takes communities into consideration, even when it is biased towards importance. This method consists of three steps.

Step 1. Apply pLSI to the citation graph to obtain $p(t|d_i, c_j)$, which is the probability that a citation from document i to another document j is made in the context of community t ($t = 1, \ldots, N$).

Create *community graph* G_t for $t = 1, \ldots, N$, with the same vertex set as the original citation graph, but assign the probability $p(t|d_i, c_j)$ as the weight to the edge from i to j. Thus the adjacency matrix of the t-th community graph $A_t = A(G_t)$ is a square matrix with (i,j)-element of $A(G_t) = p(t|d_i, c_j)$ if $(i, j) \in E$, and 0 otherwise.

Step 2. For each t, apply the Neumann kernels to the co-citation matrix $A_t^T A_t$. The Neumann kernel for community t is given by the following equation.

$$\hat{K}_{t,\gamma} = \sum_{n=1}^{\infty} (\frac{\gamma}{\lambda})^{n-1} \left(A_t^{\mathrm{T}} A_t\right)^n \tag{4}$$

This equation is identical to Eq.(1), except that A_t is used in place of the adjacency matrix of the original citation graph.

Step 3. Finally, sum the Neumann kernels in Eq. (4) over all communities t as follows.

$$R_\gamma = \sum_{t=1}^{N} \hat{K}_{t,\gamma}. \tag{5}$$

This matrix R_γ retains positive semi-definiteness, because the sum of positive semi-definite kernels is still a positive semi-definite kernel [10].

4.3 Relation to Hofmann's pLSI-Based Fisher Kernels

The Fisher kernels [11] are a method to obtain a kernel from generative models. To derive Fisher kernels from a generative model, we need to compute the Fisher score $u(d, \theta)$, the gradient of the log-likelihood function for data d with respect to the parameters θ. Given the Fisher score $u(d, \theta)$, the Fisher kernel is given by following equation.

$$K(d_i, d_j) = u(d_i; \theta)^{\mathrm{T}} I(\theta)^{-1} u(d_j; \theta)$$

Here, $I(\theta)$ is the Fisher information matrix, typically approximated by the unit matrix.

Hofmann [7] proposed a Fisher kernel for computing document similarity based on pLSI. The log-likelihood function of pLSI is given as follows.

$$\log p(d_i) = \sum_j \hat{p}(w_j|d_i) \sum_{t=1}^{N} \log p(w_j|t)p(t|d_i)$$

where $p(w_j|t)$ and $\hat{p}(w_j|d)$ respectively represent the probability that word w_j is generated from topic t, and the empirical probability of word w_j in document d. Hofmann computed the derivatives of the log-likelihood function of pLSI wrt parameters $\rho_{jt} = 2\sqrt{p(w_j|t)}$ and $\rho_t = 2\sqrt{p(t)}$ to yield two types of Fisher kernels. The derivatives wrt ρ_{jt} is given by

$$\frac{\partial \log p(d_i)}{\partial \rho_{jt}} = \frac{\hat{p}(w_j|d_i)p(t|d_i, w_j)}{\sqrt{p(w_j|t)}}.$$

So the resulting Fisher kernel is

$$\bar{K}(d, d') = \sum_{j=1} \hat{p}(w_j|d)\hat{p}(w_j|d') \sum_{t=1}^{N} \frac{p(t|d, w_j)p(t|d', w_j)}{p(w_j|t)}.$$

Analogously, we can define the Fisher kernels for citation networks by replacing word w_j with c_j, namely a citation to document j. The Fisher kernel $\bar{K}$ based on ρ_{jt} in this case can be written as a matrix

$$\bar{K} = \sum_{t=1}^{N} \bar{A}_t^{\mathrm{T}} \bar{A}_t, \tag{6}$$

where the (i,j)-element of $\bar{A}_t$ is $p(t|d_i, c_j)/\sqrt{p(c_j|t)}$ if $(i, j) \in E$, and 0 otherwise.

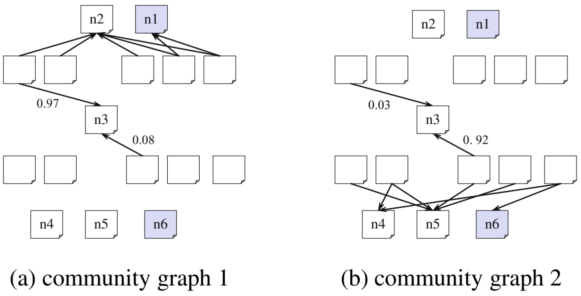
As seen from the term $\bar{A}_t^{\mathrm{T}} \bar{A}_t$ in Eq.(6), this Fisher kernel essentially computes the sum of (reweighted[2]) co-citation graphs for communities $t = 1, \ldots N$. Because computation for each community t is identical to co-citation coupling, the score is zero for any pair of documents not directly co-cited. By contrast, as mentioned in Section 4.2, our proposed kernels (Eq.(5)) compute the weighted sum of all *paths* between documents, so they assign non-zero weights to any document pair as long as they are connected in a community graph; see the discussin on the Neumann kernels in Section 2.

Hofmann used another Fisher kernel based on parameter ρ_t to compensate for this problem. This kernel is defined as

$$\tilde{K}(d, d') = \sum_{t=1}^{N} p(t|d)p(t|d')/p(t).$$

However, this kernel does not have a clear interpretation as a link analysis measure, since this equation does not involve citation c.

[2] A subtle difference between the graphs induced by $\bar{A}_t$ and A_t is that the edge weights in the former are reweighted by $\sqrt{1/p(c_j|t)}$.



(a) community graph 1 (b) community graph 2

Fig. 2. Two community graphs induced by the graph of Fig 1 (a). Edges without weight labels mean their weights equal to 1.0.

4.4 Example

The three steps of our proposed methods (Section 4.2) are demonstrated with a graph of Fig 1 (a). We set the parameter γ for Neumann kernels to 0.99, and the hyperparameter (the number of latent communities) N to 2 in pLSI. In the first step, we apply pLSI to Fig 1 (a) and obtain the two community graphs shown in Fig 2 (a), (b).

In Step 2, the Neumann kernels are applied to each community graph. Because the nodes n_1 and n_2, the most important nodes in terms of HITS authority ranking, are not connected to n_4, n_5 and n_6 in the two community graphs, the scores for n_1 and n_2 relative to the latter nodes are 0 in the respective Neumann kernels.

In Step 3, our proposed kernels (Eq. (5)) sum the Neumann kernel matrices over communities 1 and 2. As a result, the ranking for documents that only have citations within a community graph is biased towards the importance in that community. The ranking relative to the document n_3 located between the two communities is a mixture the importance in two communities.

At $\gamma = 0.99$, the final kernel $R_{0.99}$ is given as follows.

$$R_{0.99} = \begin{pmatrix} 115.04 & 235.81 & 43.56 & 0.00 & 0.00 & 0.00 \\ 235.81 & 489.90 & 91.63 & 0.00 & 0.00 & 0.00 \\ 43.56 & 91.63 & 40.82 & 37.59 & 90.96 & 10.18 \\ 0.00 & 0.00 & 37.59 & 69.38 & 157.23 & 20.07 \\ 0.00 & 0.00 & 90.96 & 157.23 & 375.79 & 42.60 \\ 0.00 & 0.00 & 10.18 & 20.07 & 42.60 & 6.70 \end{pmatrix}. \tag{7}$$

The rankings in the 4-th to 6-th rows are quite different from those of $\hat{K}_{0.99}$ (Eq. (2)). For example, in the 6-th row of $\hat{K}_{0.99}$, the highest score (2.90) is assigned to n_2, which is a node in Community 1, despite that n_6 is in Community 2. By contrast, Eq. (7) assigns the largest value (42.60) to n_5, the node with the most in-links in Community 2. We can see that our proposed kernels are biased towards the importance in the community where each document belongs.

Table 1. HITS authority rankings on community graphs. Columns T, H and P respectively show the index of the communities, HITS authority ranking in the community graph, and the ranking of pHITS.

T	H	P	Title
1	1	1	Building a large annotated corpus of English: the Penn Treebank
	2	3	Statistical decision-tree models for parsing
	3	2	A new statistical parser based on bigram lexical dependencies
	4	6	Unsupervised word sense disambiguation rivaling supervised methods
	5	5	Word-sense disambiguation using statistical models of Roget's categories trained
2	1	2	A stochastic parts program and poun phrase parser for unrestricted text
	2	1	Transformation-based error-driven learning and natural language processing
	3	3	A practical part-of-speech tagger
	4	5	A maximum entropy model for part-of-speech tagging
	5	8	MBT: A memory-based part of speech tagger-generator
3	1	3	Aligning sentences in parallel corpora
	2	1	The mathematics of statistical machine translation: parameter estimation
	3	4	Text-translation alignment
	4	5	A program for aligning sentences in bilingual corpora
	5	6	Char again: a program for aligning parallel texts at the character level
4	1	30	Generating summaries of multiple news articles
	2	25	Empirically designing and evaluating a . . . model for summary generation
	3	10	Generation of extended bilingual statistical reports
	4	33	Practical issues in automatic documentation generation
	5	20	MURAX: A robust linguistic approach for question answering . . .
5	1	1	Attention, intentions, and the structure of discourse
	2	3	Multi-paragraph segmentation of expository text
	3	4	Lexical cohesion computed by thesaural relations as an indicator of the structure of text
	4	7	Combining multiple knowledge sources for discourse segmentation
	5	11	A prosodic analysis of discourse segments in direction-giving monologues

5 Experiments

To evaluate the characteristics of the measures induced by our proposed kernels, we apply them to a bibliographic citation network in the field of natural language processing consisting of 2280 nodes (papers). In the experiments throughout this section, we set the number of communities (N) in pLSI to 5.

5.1 Community Graphs

Before examining the characteristics of the kernels, we check whether communities induced by Step 1 of our method are sensible. We computed the HITS and pHITS rankings for each community graph G_t to elucidate the central topic of each community. The top-5 lists are shown in Table 1.

From the title of the top ranked papers, we see that Community 1 represents a mixture of 'parsing' and 'word sense disambiguation'. Community 2 is the 'part-of-speech tagging' community. Community 3 is on 'machine translation'. Note that 'sentence

Table 2. The top-10 list generated by the plain Neumann kernels for 'Empirical studies in discourse' ($\gamma = 0.001$)

$\hat{K}$	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 ... 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**
11	–	8	Three generative, lexicalised models for statistical parsing

alignment' is a fundamental technique to statistical machine translation. Community 4 represents the related fields of 'summarization' and 'generation' of documents. Finally, community 5 is concerned about 'discourse processing'. There is a clear similarity between the ranking from HITS and pHITS on each community except community 4[3].

5.2 Comparison

We computed (i) plain Neumann kernels, and (ii) our proposed kernels (Eq. (5)). Each kernel matrix was treated as a ranking method by taking the i-th row vector of the matrix as the score vector for the i-th node (paper). Given the i-th score vector, or the ranking induced thereof, we call the i-th node as the *root paper* of this ranking.

To observe the differenc in the character of these kernels, We compare the top-10 papers relative to a fixed root paper, 'Empirical studies in discourse' by M. A. Walker and Johanna D. Moore, *Computational Linguistics* 23(1):1–12, 1997. This paper was used in our previous work as well, and ranks 14-th in Community 5 by HITS. Another reason we selected this root paper is that it is a paper on discourse processing. Because papers on discourse are not ranked among the top 10 HITS ranking (see column H in Table 4), the ranking for this root paper is prone to the topic drift phenomenon.

Plain Neumann Kernels. Tables 2 ($\gamma = 0.001$), 3 ($\gamma = 0.95$) and 4 ($\gamma = 0.9999$) show the list of top-10 papers induced by the plain Neumann kernels relative to the root paper 'Empirical studies in discourse.' Columns $\hat{K}$, C and H respectively display the rankings induced by the plain Neumann kernels, co-citation coupling, and the HITS authority score. A '$-$' in column C indicates that the paper was not co-cited with the root paper. Tht titles of papers on discourse processing are shown in boldface.

When $\gamma = 0.001$ (Table 2), the majority is formed by the papers on discourse processing. Ranked topmost is the root paper, followed by the six papers co-cited with the root paper, as indicated by column C.

[3] In community 4, the rankings between HITS and pHITS are not similar, but pHITS also ranked paper on summarization and generation topmost.

Table 3. The top-10 list generated by the plain Neumann kernels for 'Empirical studies in discourse' ($\gamma = 0.95$)

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

Table 4. The top-10 list generated by the plain Neumann kernels for 'Empirical studies in discourse' ($\gamma = 0.9999$)

$\hat{K}$	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
10	–	10	Integrating multiple knowledge sources to disambiguate word sense

In Table 3, we show the ranking at $\gamma = 0.95$ as a measure midway between relatedness and importance. Although the parameter value $\gamma = 0.95$ might seem too biased towards 1.0 (HITS importance), transition from relatedness to importance occurs rather late in the parameter range (typically in the range of $0.9 < \gamma < 1.0$) [3]. As a result, the ranking at $\gamma = 0.5$, for example, is mostly identical to $\gamma = 0.1$.

Only three papers on discourse remain in Table 3. These three eventually fall out of top 10 at $\gamma = 0.9999$ (Table 4). The ranking at $\gamma = 0.9999$ is identical to HITS importance ranking. We also sampled several other parameter points between $\gamma = 0.001$ and 0.9999. None of these ranking lists included discourse papers other than those appeared in Table 2 for $\gamma = 0.001$.

Proposed Kernels. Tables 5 ($\gamma = 0.001$), 6 ($\gamma = 0.95$) and 7 ($\gamma = 0.9999$) show the lists of top-10 papers induced by the proposed 'community-aware' kernels R_γ relative to 'Empirical studies in discourse.' In these tables, column R shows the rankings induced by the proposed kernels.

At $\gamma = 0.001$ (Table 5), the ranking is similar to that of the plain Neumann kernels (Table 2) and most of top ranked papers are on discourse processing.

Table 5. The top-10 list generatd by the proposed kernels for ‘Empirical studies in discourse’ ($\gamma = 0.001$)

R	C	H	Title
1	1	771	**Empirical studies in discourse**
2	2	201	**The reliability of a dialogue structure coding scheme**
3	2	76	**Assessing agreement on classification tasks: the Kappa statistic**
4	2	1061	**Effects of variable initiative on linguistic behavior in . . . language dialogue**
5	2	50	**Attention, intentions, and the structure of discourse**
6	2	1	Building a large annotated corpus of English: the Penn Treebank
7	2	604	**Message Understanding Conference (MUC) tests of discourse processing**
8	–	96	**Centering: a framework for modeling the local coherence of discourse**
9	–	374	A trainable document summarizer
10	–	60	Evaluating a focus-based approach to anaphora resolution

Table 6. The top-10 list generated by the proposed kernels for ‘Empirical studies in discourse’ ($\gamma = 0.95$)

R	C	H	Title
1	1	771	**Empirical studies in discourse**
2	2	201	**The reliability of a dialogue structure coding scheme**
3	2	76	**Assessing agreement on classification tasks: the Kappa statistic**
4	2	50	**Attention, intentions, and the structure of discourse**
5	2	1061	**Effects of variable initiative on linguistic behavior in . . . language dialogue**
6	2	1	Building a large annotated corpus of English: the Penn Treebank
7	–	96	**Centering: a framework for modeling the local coherence of discourse**
8	–	61	**Multi-paragraph segmentation of expository text**
9	–	77	**Lexical cohesion computed by thesaural relations . . .**
10	–	115	**Combining multiple knowledge sources for discourse segmentation**

Table 7. The top-10 list generated by the proposed kernels for ‘Empirical studies in discourse’ ($\gamma = 0.9999$)

R	C	H	Title
1	2	50	**Attention, intentions, and the structure of discourse**
2	–	61	**Multi-paragraph segmentation of expository text**
3	–	77	**Lexical cohesion computed by thesaural relations . . .**
4	–	115	**Combining multiple knowledge sources for discourse segmentation**
5	–	198	**A prosodic analysis of discourse segments in direction-giving monologues**
6	2	76	**Assessing agreement on classification tasks: the Kappa statistic**
7	–	150	**An automatic method of finding topic boundaries**
8	–	162	**Text segmentation based on similarity between words**
9	–	317	**Intention-based segmentation: Human reliability and correlation with . . .**
10	–	340	**Replicability of transaction and action coding in the map task corpus**

At $\gamma = 0.95$ (Table 6), we see the increase in the number of discourse papers, which is contrastive to the plain Neumann kernel with $\gamma = 0.95$ (Table 3) listing only three discource papers.

The ranking list in Table 7 for $\gamma = 0.9999$ consists solely of papers on discourse processing and text segmentation, a subtask of discourse processing. The root paper and most of the papers co-cited with the root are not on this list, but the top 5 papers match the most important papers for Community 5 (Table 1).

Fron these results, we see taht our proposed kernels did not suffer from the topic drift problem for this root paper; when γ is increased, they tend towards importance within the community where the root paper belongs. By contrast, plain Neumann kernels often assigned higher scores to papers in other communities, such as ‘A new statistical parser based on bigram lexical dependencies’; see Tables 2, 3 and 4.

6 Conclusions

We constructed a citation graph for each community using a technique similar to pLSI and pHITS. Applying Neumann kernels to each community graph, we can rank documents by taking the community of each individual document into consideration.

The technique proposed in this paper can be extended with other latent topic models beside pLSI. We are planning to apply Latent Dirichlet Allocation [12] to construct community graphs.

References

1. Kleinberg, J.M.: Authoritative sources in a hyperlinked environment. J. ACM (1999) 604–632
2. Small, H.: Co-citation in the scientific literature: a new measure of the relationship between two documents. J. American Society for Information Science **24** (1973) 265–269
3. Ito, T., Shimbo, M., Kudo, T., Matsumoto, Y.: Application of kernels to link analysis. In: Proc. 11th ACM SIGKDD. (2005) 586–592
4. Kandola, J., Shawe-Taylor, J., Cristianini, N.: Learning semantic similarity. In: NIPS 15. (2002)
5. Bharat, K., Henzinger, M.R.: Improved algorithms for topic distillation in a hyperlinked enviornment. In: Proc. 21st ACM SIGIR Conference. (1998)
6. Cohn, D., Chang, H.: Learning to probabilistically identify authoritative documents. In: Proc. 18th International Conference of Machine Learning. (2001)
7. Hofmann, T.: Learning the similarity of documents: An information-geometric approach to document retrieval and categorization. In: NIPS 12. (2000) 914–920
8. Shimbo, M., Ito, T.: Kernels as link analysis measures. In Cook, D., Holder, L., eds.: Mining Graph Data. John Wiley & Sons (2006) In press.
9. Hofmann, T.: Probabilistic latent semantic indexing. In: Proc. 22th ACM SIGIR Conference. (1999) 50–57
10. Haussler, D.: Convolution kernels on discrete structures. Technical Report UCSC-CRL-99-10, University of California at Santa Cruz (1999)
11. Jaakkola, T., Haussler, D.: Exploiting generative models in discriminative classifiers. In: NIPS 11. (1998)
12. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet Allocation. In: NIPS 14. (2001)