Efficient Image and Tag Co-Ranking: A Bregman Divergence Optimization Method

Lin Wu, Yang Wang and John Shepherd
School of Computer Science and Engineering
The University of New South Wales
Sydney, Australia
{linw,wangy,jas}@cse.unsw.edu.au

ABSTRACT

Ranking on image search has attracted considerable attentions. Many graph-based algorithms have been proposed to solve this problem. Despite their remarkable success, these approaches are restricted to their separated image networks. To improve the ranking performance, one effective strategy is to work beyond the separated image graph by leveraging fruitful information from manual semantic labeling (i.e., tags) associated with images, which leads to the technique of co-ranking images and tags, a representative method that aims to explore the reinforcing relationship between image and tag graphs. The idea of co-ranking is implemented by adopting the paradigm of random walks. However, there are two problems hidden in co-ranking remained to be open: the high computational complexity and the problem of out-of-sample. To address the challenges above, in this paper, we cast the co-ranking process into a Bregman divergence optimization framework under which we transform the original random walk into an equivalent optimal kernel matrix learning problem. Enhanced by this new formulation, we derive a novel extension to achieve a better performance for both in-sample and out-of-sample cases. Extensive experiments are conducted to demonstrate the effectiveness and efficiency of our approach.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: [Information search and retrieval]

Keywords

Co-Ranking; Bregman Divergence; Out-of-Sample

1. INTRODUCTION

The explosion of online community-contributed multimedia data results in great focus on image retrieval. Most of social media sharing websites like Flickr allow users to upload personal images and annotate content with descriptive keywords called tags. Many ranking algorithms specialized to images on such social media repositories have been proposed to help organize the shared media data [4] or to facilitate the image ranking process [7]. However, these methods mainly focus on centrality measures on image content and the evaluations of the relative importance of images have been carried out independently, which ignores the rich information from tags, and annotations, etc. Similarly, the well-studied treatments on tag recommendation [4] and tag ranking [10] are conducted solely on the tag graph. To leverage the useful metadata including tags, and images vice versa, ranking algorithms should consider the reinforcing dependency between images and tags, which is beneficial to further improving ranking results. To this end, in this paper, we study a problem of great interest, that is, how to effectively and efficiently handle dual-relational data over two graphs, which can be named as the Co-Ranking problem (CoR).

The main idea of CoR is to explore the mutually reinforcing relationship between image and tag graphs by constructing a combined graph connecting the two graphs together. Then random walks are performed to rank the data with respect to the intrinsic geometric structure revealed by a large amount of data. CoR assigns each data point a relative ranking score by collectively leveraging the ranking information from both the image graph and tag graphs. Some works have witnessed the feasibility of CoR on a variety of data types [17, 14].

One of the main drawbacks of CoR is its high computational complexity. Given a query, CoR constructs two affinity graphs and propagates the ranking scores over the combined graph, yielding to a complexity of $O(n^3)$, where n is the size of samples in the database. Such a high cost is prohibited in large-sized databases. Another limitation of CoR comes from the case of out-of-sample. Normally, if an in-sample query is issued, CoR can use off-line pre-computation to reduce the computational cost. However, if the query is out of the database, the expensive ranking score propagation step needs to be performed in the on-line stage, which is referred to be the out-of-sample problem [1].

In this paper, we reformulate the idea of random walks into a Bregman divergence optimization fashion, which provides us a novel perspective towards the co-ranking algorithm. That is, the optimal co-ranking function can be modeled by learning an optimal kernel matrix under the Bregman divergence matrix based metric. Moveover, an efficient and effective extension is induced to figure out the challenges of high computational cost and out-of-sample case.

MM'13, October 21–25, 2013, Barcelona, Spain.

http://dx.doi.org/10.1145/2502081.2502156 .

We may summarize the main contributions of this paper as follows: (i) We formulate the CoR algorithm as a Bregman divergence optimization problem. (ii) With the new formulation, we have a novel understanding of CoR's aim, that is, to learn an optimal kernel matrix, which allows a novel extension to combat the challenges of high computational cost as well as out-of-sample problem.

2. PRELIMINARIES

2.1 The Co-Ranking Algorithm

Given the well-constructed image graph G_M and tag graph G_T, the co-ranking paradigm is coupled with two random walks over G_M and G_T. Consider the transition matrix $\mathbf{Q}$ derived from G_M, and $r_k(i)$ denotes the relevance score of image v_i with respect to the query image at iteration k. The ranking score can be obtained from the unique solution as the random walk converges:

$$\mathbf{r}_\pi = (1-\rho)(\mathbf{I}-\rho\hat{\mathbf{Q}})^{-1}\mathbf{c} \tag{1}$$

where $\mathbf{c}$ is the initial ranking vector, $\mathbf{Q}$ is changed to be the ergodic $\hat{\mathbf{Q}}$ and ρ denotes the damping factor [2].

Similarly, given the transition matrix $\mathbf{Z}$ (with random walk matrix $\hat{\mathbf{Z}}$) induced by the tag graph G_T, the random walks over G_T converge to another stationary probability distribution:

$$\mathbf{f}_\pi = (1-\rho)(\mathbf{I}-\rho\hat{\mathbf{Z}})^{-1}\bar{\mathbf{c}} \tag{2}$$

where $\bar{\mathbf{c}}$ denotes the initial relevance scores of a tag.

The random walk over the combined graph is presented in terms of a random surfer who is capable of jumping over images and their tags as well. Thus, in the process of coupling two random walks, a probability distribution will have the form $(\mathbf{r}_\pi,\mathbf{f}_\pi)$ [17], satisfying $\| \mathbf{r}_\pi \|_1 + \| \mathbf{f}_\pi \|_1 = 1$.

2.2 The Bregman Matrix Divergence

Let $\phi : \Lambda \to \mathbb{R}$ be a real-valued strictly convex function defined over a convex set Λ. The Bregman divergence [3] with respect to ϕ is defined as

$$D_\phi(\mathbf{x},\mathbf{x}_0) = \phi(\mathbf{x}) - \phi(\mathbf{x}_0) - (\mathbf{x}-\mathbf{x}_0)^T\nabla\phi(\mathbf{x}_0).$$

Intuitively, the Bregman divergence is used to measure the closeness of two vectors. For example, if $\phi(\mathbf{x}) = \mathbf{x}^T\mathbf{x}$ then the corresponding Bregman divergence turns out to be the squared Euclidean distance: $D_\phi(\mathbf{x},\mathbf{x}_0) = \| \mathbf{x}-\mathbf{x}_0 \|^2$. We can naturally extend this definition to convex functions defined over matrices [8]. In this case, given a strictly convex, continuously differentiable function $\phi(\mathbf{X})$, the Bregman matrix divergence is defined to be

$$D_\phi(\mathbf{X},\mathbf{X}_0) = \phi(\mathbf{X}) - \phi(\mathbf{X}_0) - tr((\nabla\phi(\mathbf{X}_0))^T(\mathbf{X}-\mathbf{X}_0))$$

where $tr(\mathbf{X}_0)$ denotes the trace of matrix $\mathbf{X}_0$. Examples include $\phi(\mathbf{X}) = \| \mathbf{X} \|_F^2$, which leads to the squared Frobenius norm $\| \mathbf{X}-\mathbf{X}_0 \|_F^2$.

In this paper, we specialize on the log-determinant function $\phi(\mathbf{X}) = -\log det\mathbf{X}$, which can be expressed as the Burg entropy of the eigenvalues, i.e., $\phi(\mathbf{X}) = -\sum_i log\lambda_i$. The resulting matrix divergence becomes

$$\mathbb{D}(\mathbf{X},\mathbf{X}_0) = tr(\mathbf{X}\mathbf{X}_0^{-1}) - \log det(\mathbf{X}\mathbf{X}_0^{-1}) - n \tag{3}$$

which we call the Log-Determinant divergence [8].

3. BREGMAN DIVERGENCE DERIVED RANDOM WALKS

In the following, we will derive the random walks from a Bregman divergence optimization framework. By virtue of the new derivation, some extensions can be naturally derived to combat the aforementioned challenges.

Consider the convergent solution in Eq.(1), we can rewrite it as follows:

$$\mathbf{r}_\pi^* = (\mathbf{I}-\rho\hat{\mathbf{Q}})^{-1}\mathbf{c} = \mathbf{K}\mathbf{c}. \tag{4}$$

We omit the scaling factor 1-ρ as it does not influence the solutions. Let $\mathbf{T} = [\mathbf{t}_1,\mathbf{t}_2,\ldots,\mathbf{t}_n] \in \mathbb{R}^{m\times n}$ be the data representation in a new tag feature space of the data samples. In particular, we define $\mathbf{t}_i = \Phi(\mathbf{v}_i)$, for $i = 1,\ldots,n$, where Φ is a transformation function of the original image vector to the new tag feature space. Specifically, we introduce a canonical representation $[\widetilde{t}_1,\widetilde{t}_2,\ldots,\widetilde{t}_m]$ for $\mathbf{t}_i$, which is composed of the 0-1 indication of the tag feature basis. Specifically, we have

$$\mathbf{t}_i(j) = \begin{cases} 1, & \text{if } p(\mathbf{v}_i|\widetilde{t}_j) \geq \varepsilon; \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

Intuitively, we project each image into a tag feature space using a binary representation, which indicates to which extent the image is associated with basis tags.

We define the matrix $\mathbf{K}$ as

$$\mathbf{K} = \mathbf{T}^T\mathbf{T}, \tag{6}$$

which is a positive semi-definite matrix. Then, we present our primary theorem over the new derivation on random walks as follows.

Theorem 1. *The matrix $\mathbf{K}$ in the convergence formulation (Eq. (4)) is the solution of the following optimization problem:*

$$\min_{\mathbf{K}} \mathbb{D}(\mathbf{K},\mathbf{I})$$
$$s.t. \sum_{i,j} \| \tfrac{1}{d(v_i)}\mathbf{t}_i - \tfrac{1}{d(v_j)}\mathbf{t}_j \|^2 q_{ij} \leq \epsilon,\ \mathbf{K} \succeq 0,$$

where ϵ is a smoothness parameter constraining similar images having close distance in the new space.

Proof. The optimization problem seeks a $\mathbf{K}$ closest to the identity matrix measured by the Log-Determinant divergence with a regularization on normalized graph Laplacian smoothness, which can be written as a matrix form [16]:

$$\sum_{i,j} \| \frac{1}{\sqrt{d(v_i)}}\mathbf{t}_i - \frac{1}{\sqrt{d(v_j)}}\mathbf{t}_j \|^2 q_{ij} = tr(\mathbf{T}\mathbf{L}\mathbf{T}^T),$$

where $\mathbf{L} = \mathbf{I} - \hat{\mathbf{Q}}$ is the normalized graph Laplacian.

Replacing the objective function with Eq.(3) and introducing the Lagrange multiplier, the minimizing optimization problem can reformulated as follows:

$$\begin{aligned} &\min_{\mathbf{K}} \mathbb{D}(\mathbf{K},\mathbf{I}) \\ &= \min_{\mathbf{K}\succeq 0} tr(\mathbf{K}\mathbf{I}^{-1}) - \log det(\mathbf{K}) + \eta tr(\mathbf{T}\mathbf{L}\mathbf{T}^T) \\ &= \min_{\mathbf{K}\succeq 0} tr(\mathbf{K}\mathbf{I}) + \eta tr(\mathbf{T}^T\mathbf{T}\mathbf{L}) - \log det(\mathbf{K}) \\ &= \min_{\mathbf{K}\succeq 0} tr(\mathbf{K}\mathbf{E}) - \log det(\mathbf{K}), \end{aligned} \tag{7}$$

where $\mathbf{E} = \mathbf{I} + \eta\mathbf{L}$ is a positive-definite matrix and η is the Lagrange multiplier. The optimal solution $\mathbf{K}^*$ of the above optimization problem is $\mathbf{K}^* = \mathbf{E}^{-1} = (\mathbf{I} + \eta\mathbf{L})^{-1}$ [6]. Recall that $\mathbf{E} = \mathbf{I} + \eta\mathbf{L} = (1+\eta)(\mathbf{I} - \frac{\eta}{(1+\eta)}\hat{\mathbf{Q}})$, then we have $\mathbf{K}^* = (\mathbf{I} - \beta\hat{\mathbf{Q}})$, where $\beta = \frac{\eta}{(1+\eta)}$. We use $\mathbf{K}^*_M$ for the sake of distinguish. □

Similarly, we can derive another new extension from the perspective of Bregman divergence metric for the random walks over the tag graph, which is summarized in the following theorem. We define $\mathbf{K}_T$ as $\mathbf{K}_T = \mathbf{C}^T\mathbf{C}$, where $\mathbf{C} = [\mathbf{c}_1, \dots, \mathbf{c}_n] \in \mathbb{R}^{m\times n}$ is the representation that transforms the tag samples to a new feature space. Likewise, we rewrite the Eq.(2) as follows:

$$\mathbf{f}^*_\pi = (\mathbf{I} - \rho\hat{\mathbf{Z}})^{-1}\bar{\mathbf{c}} = \mathbf{K}_T\bar{\mathbf{c}}. \tag{8}$$

THEOREM 2. *The matrix $\mathbf{K}_T$ in the convergence formulation (Eq. (8)) is the solution of the following optimization problem:*

$$\min_{\mathbf{K}_T} \mathbb{D}(\mathbf{K}_T, \mathbf{I})$$
$$s.t. \sum_{i,j} \| \frac{1}{\sqrt{D_{ii}}}\mathbf{c}_i - \frac{1}{\sqrt{D_{jj}}}\mathbf{c}_j \|^2 z_{ij} \le \epsilon,\ \mathbf{K}_T \succeq 0,$$

Likewise, we have the optimal matrix $\mathbf{K}^*_T$, which is exactly equal to the matrix $\mathbf{K}_T$ in Eq. (8).

3.1 Efficient Co-Ranking Algorithm

In the perspective of Bregman divergence, the random walk essentially learns an optimal matrix $\mathbf{K}$ close to the identity matrix under certain constraints. However, we still need to inverse a $n \times n$ matrix $\mathbf{I} + \eta\mathbf{L}$, which has the complexity of $O(n^3)$. In the following, we derive an efficient extension of co-ranking algorithm. Suppose the mapping of $\mathbf{t}_i = \Phi(\mathbf{v}_i)$ is linear, i.e., $\mathbf{T} = \mathbf{P}^T\mathbf{V}$, where $\mathbf{P}$ is a $k \times m$ matrix, then we have

$$
\begin{aligned}
& tr(\mathbf{K}\mathbf{I}^{-1}) - \log det(\mathbf{K}) + \eta tr(\mathbf{T}\mathbf{L}\mathbf{T}^T) \qquad (9)\\
&= tr(\mathbf{V}^T\mathbf{P}\mathbf{P}^T\mathbf{V}) - \log det(\mathbf{V}^T\mathbf{P}\mathbf{P}^T\mathbf{V}) + \eta tr(\mathbf{P}^T\mathbf{V}\mathbf{L}\mathbf{V}^T\mathbf{P})\\
&= tr(\mathbf{H}\mathbf{V}\mathbf{V}^T) - \log det(\mathbf{V}\mathbf{V}^T) - \log det(\mathbf{H}) + \eta tr(\mathbf{H}\mathbf{V}\mathbf{L}\mathbf{V}^T),
\end{aligned}
$$

where $\mathbf{H} = \mathbf{P}\mathbf{P}^T \succeq 0$.

As a result, the optimization problem Eq. (7) becomes

$$\min_{\mathbf{H}\succeq 0} tr(\mathbf{H}(\underbrace{\mathbf{V}\mathbf{V}^T + \eta\mathbf{V}\mathbf{L}\mathbf{V}^T}_{k\times k})) - \log det(\mathbf{H}), \tag{10}$$

where $\mathbf{H}$ is a $k \times k$ matrix. To get the optimal matrix $\mathbf{H}$, we only need to inverse a matrix with size $k \times k$, which remains unchanged as the size of database (n) grows. That is, if $k \ll n$, the computational complexity of matrix inversion is reduced dramatically. Thereby, we optimize a small matrix $\mathbf{H}$ to estimate the optimal matrix $\mathbf{K}^*_T$ or $\mathbf{K}^*_M$. Then, the complexity is reduced from $O(n^3)$ to $O(m^3) + O(n^2)$.

In fact, the computation of optimal matrix $\mathbf{H}$ learns a distance metric:

$$d^2_{\mathbf{H}} = \| \mathbf{P}^T\mathbf{v}_i - \mathbf{P}^T\mathbf{v}_j \|^2 = (\mathbf{v}_i - \mathbf{v}_j)^T\mathbf{H}(\mathbf{v}_i - \mathbf{v}_j). \tag{11}$$

With the learned distance metric $\mathbf{H}$, $\mathbf{K}^*_M$ can be computed by $\mathbf{K}^*_M = \mathbf{T}^T\mathbf{T} = \mathbf{V}^T\mathbf{H}\mathbf{V}$. Hence, each element of the $\mathbf{K}^*_M$ can be calculated by

$$\mathbf{K}^*_M[ij] = \exp(-d^2_{\mathbf{H}}(\mathbf{v}_i, \mathbf{v}_j)/2\sigma^2) \tag{12}$$

Therefore, similar to Eq.(4), the ranking on image graph can use $\mathbf{K}^*_M$ to compute ranking scores:

$$\mathbf{r}^*_\pi = \mathbf{K}^*_M\mathbf{c} \tag{13}$$

3.2 Out-of-Sample Extension

Given a sample query out of the database, e.g., a new image query, we only need to compute a new column of the matrix $\mathbf{K}^*_M$, which avoids the update over the entire matrix. Assume that $\hat{\mathbf{V}} = [\mathbf{V}, \mathbf{v}_u] \in \mathbb{R}^{k\times(n+1)}$ is the new data matrix with the new sample $\mathbf{v}_u$, then we have the new optimal matrix $\hat{\mathbf{K}}^*_M \in \mathbb{R}^{(n+1)\times(n+1)}$:

$$\hat{\mathbf{K}}^*_M = \begin{pmatrix} \mathbf{K}^*_M & \mathbf{k}_{nu} \\ \mathbf{k}_{nu} & 1 \end{pmatrix} \tag{14}$$

Compared with traditional ranking algorithms based on random walks, which need $O(n^3)$ to propagate the ranking score, our approach updates the matrix $\mathbf{K}^*_M$ with the complexity of $O(n)$.

4. EXPERIMENTS

In this section, we conduct experimental studies over the task of contented-based image retrieval (CBIR) to show the effectiveness and efficiency of our approach.

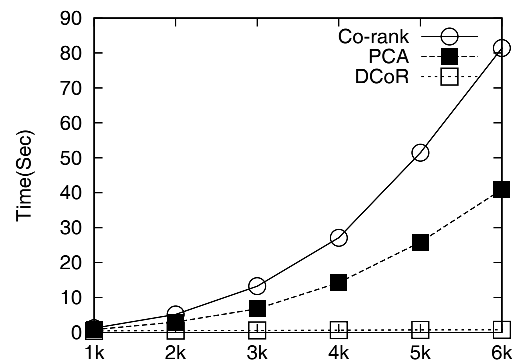

Figure 1: Time complexity. We report the running time performances of the three algorithms by ranging the size of sample from 1k to 6k. This experiment is performed over the benchmark of Caltech101 database.

4.1 Experimental Setup

Databases. All experiments are conducted on two real-world image databases: COREL and Caltech101. COREL is a widely used database for CBIR, which contains 7,000 images categorized into 70 classes [15, 5]. The Caltech101 database contains about 9,000 images divided into 101 categories with 90 averaged images per category [1].

Baselines. We implemented the following algorithms as baselines.

- PCA: PCA is the most popular linearly dimensional reduction method [13]. We set the reduced dimensionality to be 40 in our experiments.
- CoR: Co-Ranking is the simple paradigm using combined random walks to do image and tag ranking [14].
- DCoR: the efficient extension on co-ranking made in this paper by using the Bregman divergence principle.

[1] $http : //www.vision.caltech.edu/Image_Datasets/Caltech101$

Image Features. We extract three features from each image: Grid color moment, wavelet texture [9] and local binary pattern [12]. Thereby, a 260-dimensional vector is formed to describe images.

Evaluation Metrics. Three evaluation metrics are considered: the Precision at top $\mathcal{K}$ ($\mathcal{P}@\mathcal{K}$), NDCG@$\mathcal{K}$ and Mean Average Precision (MAP) [11]. We present the definitions of NDCG@$\mathcal{K}$ and MAP in the follows:

$$NDCG@\mathcal{K} = Z_n \sum_{j=1}^{\mathcal{K}} \frac{2^{r(j)} - 1}{\log_2(j+1)}, \tag{15}$$

where $r(j)$ is the relevance score of the jth returned item and Z_n is a normalization constant.

MAP, as defined in [11] is:

$$MAP(Q) = \frac{1}{|Q|} \sum_{j=1}^{|Q|} \frac{1}{m_j} \sum_{k=1}^{m_j} Precision(R_{jk}) \tag{16}$$

where Q is a set of queries, q_j is one query from Q with m_j relevant items $\{i_1, \ldots, i_{m_j}\}$, and R_{jk} is the set of ranked items for q_j until i_k is reached.

4.2 Performance Evaluations

Table 1: Results on COREL and Caltech101 data sets over $\mathcal{P}@\mathcal{K}$, NDCG@$\mathcal{K}$ and MAP.

	COREL			Caltech101		
Metric	PCA	CoR	DCoR	PCA	CoR	DCoR
$\mathcal{P}$@10	40.57%	42.99%	**44.77%**	34.54%	34.94%	**38.65%**
$\mathcal{P}$@20	34.46%	37.55%	**38.35%**	29.83%	30.87%	**35.79%**
$\mathcal{P}$@30	31.83%	**34.67%**	34.05%	29.83%	30.88%	**33.73%**
NDCG@10	42.49%	45.09%	**47.51%**	35.43%	36.31%	**40.37%**
NDCG@20	38.12%	41.02%	**41.98%**	33.15%	36.07%	**37.67%**
NDCG@30	34.47%	38.20%	**38.27%**	31.81%	32.84%	**36.05%**
MAP	32.38%	**36.71%**	35.47%	31.07%	35.22%	**36.56%**

Ranking evaluations. As shown in Table 1, in terms of COREL database, as the image features are discriminative, the baselines of PCA and CoR perform reasonably well by simply using the Euclidean distance. Our approach only loses slightly to the best performance shown by CoR.

For the Caltech101 database, the method of PCA doesn't show good performance, partially due to the large number of categories and complex content in images. Also, the approach of CoR is unable to show promising results because the graphs used in CoR are based on the Euclidean distance, which cannot capture the true metrics between images. In contrast, our method constantly shows superior results over baselines by the virtue of distance metric learning.

Time complexity. We show running time complexity in Fig.1. It can be seen that CoR needs high computational cost and it is hard to be applied on the application of large-scaled databases. Although PCA reduces the running time to some degree, the cost of its computation is still not low enough to make it applicable on large-sized data sets. In contrast, our approach is more efficient as the size of database increases.

5. CONCLUSIONS

In this paper, we present a new perspective of random walks, which learns an optimal kernel matrix implemented by the Bregman matrix divergence metric. This novel formulation allows us to derive an efficient extension over the co-ranking algorithm as well as an effective strategy for the case of out-of-sample. The two extensions reduce the computational complexity dramatically. Extensive experiments are conducted to show the superiority of our approach.

6. REFERENCES

[1] Y. Bengio, J. Paiement, P. Vincent, O. Delalleau, N. L. Roux, and M. Ouimet. Out-of-sample extensions for lle, isomap, mds, eigenmaps and spectral clustering. In *NIPS*, pages 177–184, 2004.

[2] P. Berkhin. A survey on pagerank computing. *Internet Mathematics*, 2(1):73–120, 2005.

[3] L. Bregman. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR computational mathematics and mathematical physics*, 7(3):200–217, 1967.

[4] Z. Guan, J. Bu, Q. Mei, C. Chen, and C. Wang. Personalized tag recommendation using graph-based ranking on multi-type interrelated objects. In *SIGIR*, 2009.

[5] X. He, D. Cai, and J. Han. Learning a maximum margin subspace for image retrieval. *IEEE Transactions on Knowledge and Data Engineering*, 20(2):189–201, 2007.

[6] S. C. H. Hoi, W. Liu, and S.-F. Chang. Semi-supervised distance learning for collaborative image retrieval and clustering. *ACM Transactions on multimedia computing, communications and applications*, 6(3), 2010.

[7] Y. Jing and S. Baluja. Pagerank for product image search. In *WWW*, 2008.

[8] B. Kulis, M. Sustik, and I. Dhillon. Learing low-rank kernel matrices. In *ICML*, 2006.

[9] M. Lades, J. Vorbruggen, J. Buhmann, J. Lange, C. von der Malsburg, R. Wurtz, and W. Konen. Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions on Computers*, 42(3):300–311, 1993.

[10] D. Liu, X.-S. Hua, L. Yang, M. Wang, and H.-J. Zhang. Tag ranking. In *WWW*, pages 351–360, 2009.

[11] C. Manning, P. Raghavan, and H. Schutze. *Introduction to information retrieval*. Cambridge University Press, Cambridge, 2008.

[12] T. Ojala, M. Pietikainen, and D. Harwood. A comparative study of texture measures with classification based on featured distributions. *Pattern recognition*, 29(1):51–59, 1996.

[13] J. Shlens. A tutorial on principal component analysis. Technical report, Measurement, 2005.

[14] L. Wu, Y. Wang, and J. Shepherd. Co-ranking images and tags via random walks on a heterogeneous graph. In *Multimedia Modeling*, 2013.

[15] B. Xu, J. Bu, C. Chen, D. Cai, X. He, W. Liu, and J. Luo. Efficient manifold ranking for image retrieval. In *SIGIR*, pages 525–534, 2011.

[16] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Schölkopf. Learning with local and global consistency. In *NIPS*, 2003.

[17] D. Zhou, S. A. Orshanskiy, H. Zha, and C. L. Giles. Co-ranking authors and documents in a heterogeneous network. In *ICDM*, pages 739–744, 2007.