of a document, which is generally set to be uniform. Indeed, $p(d)$ can also be set to be how popular the d is (a variant of the citation count, much like a PageRank [3] for papers). In many situations, this may improve the expert retrieval. In this paper, we followed the model proposed by Deng et al. [7] to estimate $p(d)$ using the natural logarithm of the citation count c_d of the document: $p(d) \propto \ln(e + c_d)$, which achieved much better results than the uniform setting.

For simplicity, let $\mathbf{x}$ be the relevance vector with $x_i = p(q|d_i)$, $\mathbf{y}$ be the expertise vector with $y_j = p(a_j|q)$, and $\mathbf{Q}_D \in \mathbb{R}^{|D|\times|D|}$ be a diagonal matrix with $\mathbf{Q}_{D_{ii}} = p(d_i)$. The *baseline* model as shown in Eq. (1) can be rewritten as:

$$\mathbf{y} = \mathbf{P}_{DA}^T \mathbf{Q}_D \mathbf{x}, \tag{2}$$

where $\mathbf{P}_{DA} \in \mathbb{R}^{|D|\times|A|}$ is the transition matrix from documents $\mathcal{D}$ to authors $\mathcal{A}$ with $\mathbf{P}_{DA_{ij}} = p(a_j|d_i)$. The underlying intuition of this baseline model is to *estimate the expertise of a candidate based on the relevance and quality of associated documents*, which is illustrated as the left up-arrow in Figure 1(b).

2.2 Preliminaries

In this subsection, we define the problem of expertise ranking in a heterogeneous network and introduce several related concepts and notations.

A bibliographic heterogeneous network consists of three types of object sets, including an author set $\mathcal{A} = \{a_1, a_2, ..., a_m\}$, a document set $\mathcal{D} = \{d_1, d_2, ..., d_n\}$ and a venue set $\mathcal{C} = \{c_1, c_2, ..., c_l\}$, as well as the textual content information for each document. Such a heterogeneous network of authors, documents and venues can be denoted as $G = (V, E)$ where $V = V_A \cup V_D \cup V_C$ and $E = E_A \cup E_D \cup E_{A,D} \cup E_{D,C}$. As mentioned before, different types of edges should be treated differently, thus three kinds of graphs are formed as shown in Figure 2. G_D is an unweighted directed graph of documents, G_A is an undirected graph (co-authorship graph or social network) of authors, while $G_{A,D,C}$ is the tripartite graph representing authorship and publishing relationships.

Basically, G_D can be constructed based on a directed citation graph of papers or an affinity graph of documents. With respect to citation graph, a unidirectional link is formed from d_i to $d_j (i \neq j)$, denoted by $\mathbf{W}_{ij} = 1$, if document d_i cites document d_j, and otherwise no link is constructed. Thus, we construct a directed graph G_D with adjacency matrix $\mathbf{W} \in \mathbb{R}^{|D|\times|D|}$. For a given bibliography dataset (e.g., DBLP bibliography), it is very easy to obtain the bipartite graph $G_{D,A}$ and its matrix $\mathbf{R} \in \mathbb{R}^{|D|\times|A|}$, where $\mathbf{R}_{ij} = 1$ if d_i is associated with a_j. In addition, we use $\mathbf{P}_{DA} \in \mathbb{R}^{|D|\times|A|}$ to denote the transition matrix from documents to authors, where $\mathbf{P}_{DA} = \mathbf{D}_R^{-1}\mathbf{R}$ and $\mathbf{D}_R \in \mathbb{R}^{|D|\times|D|}$ is a diagonal matrix with $\mathbf{D}_{R_{ii}} = \sum_j R_{ij}$, thus it satisfies $\mathbf{P}_{DA_{ij}} = p(a_j|d_i)$. Similarly, $\mathbf{P}_{AD} \in \mathbb{R}^{|A|\times|D|}$ is defined as the transition matrix from authors to documents with $\mathbf{P}_{AD} = \mathbf{D}_{R^T}^{-1}\mathbf{R}^T$.

Based on $G_{D,A}$, we can construct the co-authorship graph G_A as well as its corresponding matrix $\mathbf{A}$. To quantify the edge weight, we chose the co-authorship frequency [15] that is the sum of values for all papers co-authored by a_i and a_j, $\mathbf{A}_{ij} = \sum_{k=1}^{N} \frac{\delta_i^k \delta_j^k}{n_{d_k}-1}$, where $\delta_i^k = 1$ if a_i is one of the authors of paper d_k, $\delta_i^k = 0$ otherwise, and n_{d_k} is the number of authors in paper d_k. Figure 2(b) illustrates the weighted co-authorship graph of the example as shown in Figure 2(c). In this way, the more the co-authored papers, the higher the edge weight between two persons.

Now we can formulate our *expertise ranking* problem as: Given a heterogeneous network G, and the textual content information of documents $\mathcal{D}$, for a query q, we want to identify the relevant and knowledgable experts with expertise in a particular field. Previous work has explored the textual content information of documents $\mathcal{D}$ for ranking expertise based on document-centric model. In the following section, we investigate how to enhance the expertise ranking performance by modeling and exploiting the heterogeneous network G on top of the baseline model.

3. MODELING HETEROGENEOUS NETWORKS

Generally, documents and authors are closely connected in the heterogeneous bibliographic network, and much more information is available in addition to the textual document information. In this section we discuss several ways of incorporating different types of graphs into expertise ranking by developing regularization constraints.

3.1 Basic Hypotheses

We now describe our hypotheses to incorporate the heterogeneous bibliographic networks, and investigate how these hypotheses can be applied in enhancing the performance of expertise ranking. In Section 4 we validate them experimentally on real-world data, and demonstrate that they hold for the data sets we consider, which can also be generalized to other domains with similar semantics. Based on common sense and our observations on real data, we have the following three basic hypotheses:

Document Consistency Hypothesis: *Usually similar documents will be of similar relevance to a given query.* In a citation (or affinity) graph, it is reasonable to assume the neighbors of a document d are those documents that are considered similar to it. If the neighbors of a document are highly relevant to a query, this document is more likely to be relevant to the query; otherwise, if none of the neighbors of a document is relevant to the query, the document is unlikely to be relevant to the query.

Co-Authorship Consistency Hypothesis: *If two persons have co-authored many relevant papers with respect to a given query, then their expertise in the queried field should be similar in some sense.* Suppose author a_1 is an expert in a specific research topic, e.g., "information retrieval", and has co-published many papers related to this topic with author a_2, then author a_2 is more likely to be knowledgeable about this topic.

Document-Author Consistency Hypothesis: *The expertise of a researcher is consistent with that of the documents he/she published.* In this case, author a_i is knowledgeable about a specific topic only if author a_i has published papers related to the topic. Similarly, paper d_j may be related to the specific topic if paper d_j is written by authors who are experts in this area.

3.2 Regularization Framework

We now describe how we enforce the above hypotheses by defining the regularization constraints for the expertise ranking model.

3.2.1 Document Consistency

Here the goal is to refine the relevance score vector $\mathbf{x}$ based on the document consistency that *similar documents are likely to have the same ranking scores for a given query.* Suppose we are given a document citation graph $G_D = (V_D, E_D)$ as illustrated in Figure 2(a), which is a directed graph. Suppose the pairwise similarities among the documents are described by the matrix $\mathbf{S}_D \in \mathbb{R}^{|D|\times|D|}$ measured based on G_D. In the baseline model, the relevance of documents is calculated using the language model for a given query, denoted by the initial relevance vector $\mathbf{x}^0$. Thus we formulate to minimize a regularization loss function [29] as follows:

$$\Omega_1 = \mathbf{x}^T(I - \mathbf{S}_D)\mathbf{x} + \mu_d \left\|\mathbf{x} - \mathbf{x}^0\right\|^2, \quad (3)$$

where $\mu_d > 0$ is the regularization parameter. The first term of the cost function defines the *document consistency*, which prefers small difference in relevance scores between neighbor documents, while the second term defines the *fitting constraint* that measures the difference between the final scores $\mathbf{x}$ and the initial relevance scores $\mathbf{x}^0$.

Minimizing Ω_1 will force the neighbor documents to receive similar relevance scores. Differentiating Eq. (3) and setting $\partial\Omega_1/\partial\mathbf{x} = 0$, we can see that the optimal solution $\mathbf{x}^*$ may be written as

$$\mathbf{x}^* = (1-\alpha)(I - \alpha\mathbf{S}_D)^{-1}\mathbf{x}^0, \quad (4)$$

where $\alpha = 1/(1+\mu_d)$. Accordingly, it needs to calculate the inverse matrix $(I - \alpha\mathbf{S}_D)^{-1}$ to get the optimal solution. Fortunately, the matrix $\mathbf{S}_D$ is usually very sparse, then the complexity time of the sparse matrix inversion can be reduced to be linear with the number of nonzero matrix elements. One alternative solution to the above can be obtained using a powerful iterative method [29, 31]: $\mathbf{x}(t+1) = \alpha\mathbf{S}_D\mathbf{x}(t) + (1-\alpha)\mathbf{x}^0$, where $\mathbf{x}^* = \mathbf{x}(\infty)$ is the solution. In general, the iterative algorithm can converge after 10 iterations in most cases, which means the regularization problem can be solved efficiently with the iterative method. Then the expertise vector $\mathbf{y}^*$ can be inferred according to Eq. (2), so as to affect the results of expertise ranking.

Now the interesting question is how to calculate $\mathbf{S}_D$ among the set $\mathcal{D}$. For graph data, a number of recent work [31] has been given on obtaining the similarity measures. For a directed graph G_D, where the adjacency matrix $\mathbf{W}$ is first normalized as a random walk transition matrix $\mathbf{P}_D = \mathbf{D}_W^{-1}\mathbf{W}$, the similarity measure $\mathbf{S}_D$ is calculated as:

$$\mathbf{S}_D = \frac{\Pi_D^{1/2}\mathbf{P}_D\Pi_D^{-1/2} + \Pi_D^{-1/2}\mathbf{P}_D^T\Pi_D^{1/2}}{2},$$

where Π_D is a diagonal matrix formed from the stationary probability distribution of the adjacency matrix, and $\mathbf{D}_W$ is a diagonal matrix with $\mathbf{D}_{D_{ii}} = \sum_j \mathbf{W}_{ij}$. For undirected graph, $\mathbf{S}_D$ is simply the normalized adjacency matrix: $\mathbf{S}_D = \mathbf{D}_W^{-1/2}\mathbf{W}\mathbf{D}_W^{-1/2}$. In this paper, we consider two different document graphs: One is the citation graph which is a directed graph as shown in Figure 2(a), and the other is the co-conference graph which is an undirected graph with $\mathbf{W}_{ij} = 1$ if d_i and d_j appear in a same conference.

3.2.2 Co-Authorship Consistency

The objective of co-authorship consistency is to *enforce the expertise scores of candidates to be closer if they co-authored more papers related to a given query.* Suppose we are given the co-authorship graph $G_A = (V_A, E_A)$, which is a weighted undirected graph. Let $\mathbf{A} \in \mathbb{R}^{|A|\times|A|}$ be the co-authorship matrix based on G_A. The regularization for the co-authorship consistency is very similar to the one for the document consistency, which can be defined as

$$\Omega_2 = \mathbf{y}^T(I - \mathbf{S}_A)\mathbf{y} + \mu_a \left\|\mathbf{y} - \mathbf{y}^0\right\|^2, \quad (5)$$

$$s.t. \quad \mathbf{y}^0 = \mathbf{P}_{DA}^T\mathbf{Q}_D\mathbf{x}, \quad (6)$$

where $\mathbf{S}_A = \mathbf{D}_A^{-1/2}\mathbf{A}\mathbf{D}_A^{-1/2}$ is the normalized matrix, $\mathbf{x}$ can either be the initial relevance scores $\mathbf{x}^0$, or the optimal solution $\mathbf{x}^*$ refined by document consistency. Intuitively, the first term of Eq. (5) defines the *co-authorship consistency*, which will push the expertise of an author a_i to be close to his/her co-authors if they published many relevant papers together, while the second term is the constraint to fit the expertise scores obtained by the baseline (or refined) model.

The solution of minimizing Ω_2 can be achieved with the closed form solution:

$$\mathbf{y}^* = (1-\beta)(I - \beta\mathbf{S}_\mathbf{A})^{-1}\mathbf{y}^0,$$
$$= (1-\beta)(I - \beta\mathbf{S}_\mathbf{A})^{-1}\mathbf{P}_{DA}^T\mathbf{Q}_D\mathbf{x}, \quad (7)$$

where $\beta = 1/(1+\mu_a)$. Similarly, the iterative solution becomes: $\mathbf{y}(t+1) = \beta\mathbf{S}_A\mathbf{y}(t) + (1-\beta)\mathbf{P}_{DA}^T\mathbf{Q}_D\mathbf{x}$, where $\mathbf{y}^* = \mathbf{y}(\infty)$ is the solution. By setting $\mathbf{x} = \mathbf{x}^*$, the above optimization problem can be interpreted as considering the document consistency and co-authorship consistency sequentially in a two-stage process. Although the co-authorship consistency is derived by considering the co-authorship graph, some other information, for example, co-worker information in a same department, can be transformed into a co-authorship graph to some extent, which means our model can handle more general information sources instead of the co-authorship.

3.2.3 Document-Author Consistency and Joint Regularization Framework

The key point of the document-author consistency is that *the expertise of an author is consistent with the relevance of associated documents.* In the above, we have given the methods for incorporating the document consistency and the co-authorship consistency into expertise ranking, respectively. Actually, the document-author consistency has been explicitly used in these methods as well as the baseline model, according to Eq. (2). It can be viewed as a propagation from documents (relevance) to authors (expertise) based on the transition matrix $\mathbf{P}_{DA}$ from $\mathcal{D}$ to $\mathcal{A}$.

In contrast, it is essential to investigate whether the expertise of authors could reinforce the relevance of their associated documents with respect to a query, based on another transition matrix $\mathbf{P}_{AD}$ from $\mathcal{A}$ to $\mathcal{D}$. The underlying assumption is that the expertise/knowledge of an author could propagate to the associated documents according to $\mathbf{Q}_D\mathbf{x} = \mathbf{P}_{AD}^T\mathbf{y}$. Therefore, we can encode the support from the associated authors with the initial relevance scores, so as to define the following new value:

$$\hat{\mathbf{x}}^0 = (1-\gamma)\mathbf{x}^0 + \gamma\mathbf{Q}_D^{-1}\mathbf{P}_{AD}^T\mathbf{y}, \quad (8)$$

where γ is the parameter to control the balance between the initial relevance scores and the propagated scores. Note that *if $\gamma = 0$ we only consider the propagation from documents to authors, while ignore the propagation from authors to documents.* When $\gamma > 0$, we take both propagations into account, which can be denoted as *Mutual Document-Author Consistency.*

To incorporate all the three hypotheses on a heterogeneous network, formally, a joint objective function is defined to be

$$\begin{aligned}\Omega_3 &= \Omega_1 + \Omega_2, \\ &= \mathbf{x}^T(I - \mathbf{S}_D)\mathbf{x} + \mu_d \left\|\mathbf{x} - \hat{\mathbf{x}}^0\right\|^2 \\ &\quad + \mathbf{y}^T(I - \mathbf{S}_A)\mathbf{y} + \mu_a \left\|\mathbf{y} - \mathbf{y}^0\right\|^2. \end{aligned} \quad (9)$$

along with the constraints as defined in Eq. (6) and Eq. (8). One can understand the above optimization problem in this way: Here Ω_1 is responsible for the *document consistency* within documents, while Ω_2 is responsible for the *co-authorship consistency* within authors. In the meanwhile, the constraints can be considered as the *document-author consistency* between documents and authors.

The optimization illustrated above can be solved using the standard conjugate gradient method, and a closed-form solution can be derived (The proof is omitted due to space limit). However, for a large-scale information retrieval, an iterative algorithm would be more effective and preferable to solve the optimization problem. Suppose $\mathbf{x}(0) = \mathbf{x}^0$ and $\mathbf{y}(0) = \mathbf{P}_{DA}^T\mathbf{Q}_D\mathbf{x}^0$, in the $(t+1)$-th iteration, we first compute $\mathbf{x}(t+1)$ using $\mathbf{x}(t)$ and $\mathbf{y}(t)$:

$$\begin{aligned}\mathbf{x}(t+1) &= \alpha\mathbf{S}_D\mathbf{x}(t) + (1-\alpha)\left((1-\gamma)\mathbf{x}^0 + \gamma\mathbf{Q}_D^{-1}\mathbf{P}_{AD}^T\mathbf{y}(t)\right) \\ &\quad + \alpha\frac{1-\beta}{\beta}\mathbf{Q}_D\mathbf{P}_{DA}\left(\mathbf{y}(t) - \mathbf{P}_{DA}^T\mathbf{Q}_D\mathbf{x}(t)\right), \end{aligned} \quad (10)$$

and then compute $\mathbf{y}(t+1)$ based on $\mathbf{x}(t+1)$ and $\mathbf{y}(t)$:

$$\begin{aligned}\mathbf{y}(t+1) &= \beta\mathbf{S}_A\mathbf{y}(t) + (1-\beta)\mathbf{P}_{DA}^T\mathbf{Q}_D\mathbf{x}(t+1) \\ &\quad + \beta\frac{1-\alpha}{\alpha}\mathbf{P}_{AD}\mathbf{Q}_D^{-1}\left(\mathbf{x}(t+1) - (1-\gamma)\mathbf{x}^0 - \gamma\mathbf{Q}_D^{-1}\mathbf{P}_{AD}^T\mathbf{y}(t)\right). \end{aligned} \quad (11)$$

Essentially, the values of $\mathbf{x}$ and $\mathbf{y}$ will be updated iteratively according to the parameter setting and the heterogeneous graphs until they converge. The trade-off among these hypotheses is controlled by the parameters $\alpha = 1/(1+\mu_d)$, $\beta = 1/(1+\mu_a)$ and γ, where each of them ranges from 0 to 1. In this paper, we employ a grid-search on the parameters α, β and γ using cross-validation [12], and report the performance in Section 4.

3.3 Connections and Discussions

Here we establish connections between the joint regularization framework and other methods in Table 1. Suppose $\alpha = 0$ ($\mu_d \to \infty$), $\beta = 0$ ($\mu_a \to \infty$), $\gamma = 0$, as shown in Eq. (9), Ω_1 puts all weight $\mu_d \to \infty$ on the second term $\left\|\mathbf{x} - \hat{\mathbf{x}}^0\right\|^2$ to ensure that the final scores are equal to the initial scores $\mathbf{x}^0$ and then propagate to the candidates. Similarly, Ω_2 puts all weight $\mu_a \to \infty$ on the term $\left\|\mathbf{y} - \mathbf{y}^0\right\|^2$, so as to ensure that the final scores are equal to $\mathbf{y}^0$. In this case, the optimal results are $\langle \mathbf{x}^0, \mathbf{P}_{DA}^T\mathbf{Q}_D\mathbf{x}^0 \rangle$, and the regularization framework boils down to the *baseline* model.

To incorporate the consistency of each of the three hypothesis individually, it is equivalent to explore one of the parameters α, β, γ and fix the other two parameters to 0 as shown in Table 1. For example, by setting $\alpha > 0$, $\beta = \gamma = 0$, Eq. (11) becomes $\mathbf{y}(t+1) = \mathbf{P}_{DA}^T\mathbf{Q}_D\mathbf{x}(t+1)$, and then Eq. (10) is simplified to the iterative solution of Eq. (3), which means the model only considers the *document consistency* in the regularization framework. Suppose $\gamma > 0$, $\alpha = \beta = 0$, the model consider the *mutual document-author consistency*, and the iterative solution becomes:

$$\mathbf{x}(t+1) = (1-\gamma)\mathbf{x}^0 + \gamma\mathbf{Q}_D^{-1}\mathbf{P}_{AD}^T\mathbf{y}(t),\ \mathbf{y}(t+1) = \mathbf{P}_{DA}^T\mathbf{Q}_D\mathbf{x}(t+1),$$

which is similar to HITS algorithm [14]. Finally, we will investigate the general case (i.e., $\alpha > 0$, $\beta > 0$, $\gamma > 0$) which combines multiple consistency hypotheses jointly.

To compute the model efficiently, we only need to retrieve a subset of documents $\hat{D}$ with top-k relevant documents, and then identify the subset of authors $\hat{A}$ associated with $\hat{D}$, so as to perform our model on the subgraph with most relevant documents and authors. We can benefit from this trick to reduce computational cost and enhance the accuracy by avoiding topic drift in the whole graph. However, the selection of k can also be an issue, which is empirically studied in Section 4.2.4. In addition, it is worth mentioning that the key contribution of this paper is to exploit heterogeneous networks for enhancing expertise ranking model, and our model is built on top of the baseline model.

Table 1: Connections with other methods.

Parameters: α, β, γ	Description
$\alpha = 0,\ \beta = 0,\ \gamma = 0$	*Baseline* probabilistic model
$\alpha > 0,\ \beta = 0,\ \gamma = 0$	Document consistency[a]
$\alpha = 0,\ \beta > 0,\ \gamma = 0$	Co-authorship consistency[a]
$\alpha = 0,\ \beta = 0,\ \gamma > 0$	Mutual document-author consistency
$\alpha > 0,\ \beta > 0,\ \gamma > 0$	Joint Regularization

[a] Note that document-author consistency is also partially used in these models as described in Section 3.2.3.

Table 2: Statistics of the DBLP network.

Nodes		Edges	
# of authors	695,906	# edges in G_A	4,272,319
# of papers	1,152,512	# edges in G_D	5,695,135
# of venues	3,311	# edges in $G_{D,A}$	2,944,797

4. EXPERIMENTS

4.1 Experimental Setup and Metrics

4.1.1 Data Collection

We evaluate our models on the real-world DBLP bibliography data[1], which contains over 1,100,000 XML records. Each record consists of several elements, such as "author", "title", "conference". Using these records, we could easily build the paper-author bipartite graph $G_{D,A}$, the paper-venue bipartite graph $G_{D,C}$ and the co-authorship graph G_A. The citation graph G_D and abstract information were obtained from Arnetminer[2]. In addition, we can construct the document co-conference graph $G'_{\mathcal{D}}$ according to the bipartite graph $G_{D,C}$ for another alternative instead of citation graph.

From the statistics in Table 2, we get totally 1,152,512 papers and 695,906 authors in our data collection. After the construction of the heterogeneous graph, we observe that there are relatively few edges, indicating the resulting matrices are sparse matrices. As for G_A, each author has 6.13 co-authors on average. In G_D we only get about 5 citations per paper. This is because the citations of some papers may not be available in the repository. It can be imagined that we could achieve better results with the whole citation graph, but it is enough to illustrate the performance of our

[1] http://www.informatik.uni-trier.de/~ley/db/
[2] http://arnetminer.org/lab-datasets/citation/

Table 3: Benchmark dataset of 20 queries.

Topic	#Expert
Information Extraction	20
Intelligent Agents	29
Machine Learning	42
Natural Language Processing	41
Planning	34
Semantic Web	45
Support Vector Machine	31
Boosting	56
Ontology Alignment	53
Probabilistic Relevance Model	13
Information Retrieval	23
Language Model for Information Retrieval	12
Face Recognition	21
Semi Supervised Learning	21
Reinforcement Learning	17
Kernel Methods	21
Privacy Preservation	17
Skyline	12
Sensor RFID data management	13
Stream	16

proposed model using the current citation graph. With regard to $G_{D,A}$, on average there are 4.22 papers per author, and 2.55 authors per paper. Since these matrices are very sparse but they represent real-world situations, our proposed algorithm can efficiently solve for the results.

The evaluation of expert finding performance in such a large data collection is very challenging due to the scarcity of ground truth that can be examined publicly. Furthermore, it is impractical to obtain expert ratings for all authors. In order to measure the performance of our proposed methods, a benchmark dataset with 20 query topics and expert lists is manually created as shown in Table 3. The top 9 topics and expert lists in the left table were collected by Zhang et al. [28, 7], which are available at `http://keg.cs.tsinghua.edu.cn/project/PSN/dataset.html#new_expert_list`. The rest 11 topics and relevance judgments for the corresponding expert lists were created and evaluated by 10 researchers and senior graduate students of CUHK [8]. Specifically, these 11 topics were created by the assessors based on their own research topics. We tried to cover not only broad queries but also specific queries to see whether our methods could handle both of them effectively.

Following general relevance judgments, for each query, a list of relevant experts is collected through the method of pooled relevance judgments with human assessment efforts. The top ranked/retrieved authors from the computer science bibliography search engines (such as CiteSeer, Libra, Rexa and ArnetMiner), and the committees of the top conferences related the query topic were taken to construct the pools which contained around 300 authors. Moreover, since the query topics were created by the assessors who had conducted research in the related field for several years, they were quite familiar with the experts in that research field and could make reliable relevance judgments and even nominate some missing experts. The assessments were carried out mainly in terms of the number of top conference/journal papers an expert candidate had published, the number of related publications for the given query, and what distinguished awards he/she had received. There are four grade scores (3, 2, 1, and 0) which were assigned respectively to represent top expert, expert, marginal expert, and non-expert. Basically, each query and the corresponding expert list are judged by at least 3 (to 5) assessors, and we intentionally obtained a small number of experts by marking around 20 top ranked experts as top experts (although the number of experts could be quite large). Finally, the judgment scores (at levels 3 and 2) were averaged to construct the final ground truth with 20 to 50 experts for each query as shown in Table 3. This dataset based on DBLP [8] has been widely used for evaluating expert finding.

4.1.2 Evaluation Metrics

For the evaluation of the task, several popular IR metrics are employed to measure the performance of our proposed models, including precision at rank n ($P@n$), Mean Average Precision (MAP), bpref [4, 24], and Mean Reciprocal Rank (MRR). $P@n$ measures the fraction of the top-n retrieved results that are relevant experts for the given query, which is defined as $P@n = \frac{\#\text{ relevant in top } n \text{ results}}{n}$. R-precision (R-prec) is defined as the precision at rank R where R is the number of relevant candidates for the given query. Average precision (AP) emphasizes returning more relevant documents earlier. For a single query, AP is defined as the average of the $P@n$ values for all relevant documents: $AP = \frac{\sum_{n=1}^{N}(P@n * \text{rel}_{(n)})}{R}$, where n is the rank, N the number retrieved, and $\text{rel}(n)$ is a binary function indicating the relevance of a given rank. MAP is the mean value of the average precisions computed for all queries. The Mean Reciprocal Rank(MRR) of each individual query is the reciprocal of the rank at which the first relevant answer was returned.

Beside the measurement of precisions, Bpref [4] is a good score function that evaluates the performance from a different view, i.e., the number of non-relevant candidates. It is formulated as $\text{bpref} = \frac{1}{R}\sum_{r=1}^{N}(1 - \frac{\#n \text{ ranked higher than } r}{R})$, where r is a relevant candidate and n is a member of the first R candidates judged non-relevant as retrieved by the system.

4.2 Experimental Results

In this subsection, we evaluate our proposed model for expertise ranking. First, the experiments are performed to compare with several state-of-the-art methods, and also to validate our model under various hypotheses. Then we investigate the parameter effect of these hypotheses. Finally, we analyze the detailed results with some case studies.

4.2.1 Comparison of Different Methods

To demonstrate how the expertise ranking performance can be improved by our proposed approaches, we implemented several state-of-the-art methods as follows:

Balog's Model2 [1] is a popular document-centric model for expert finding. It is one of the best-performing models by only utilizing the documents. The source code is available at `http://code.google.com/p/ears`.

BL [7, 8] is our baseline probabilistic model as described in Section 2.1. This model estimates the expertise of a candidate based on both the relevance and quality of associated documents.

EnhancedBL [8] is an enhanced model for expertise ranking with two community-aware strategies. More specifically, it incorporates the co-authorship graph with query-

Table 4: Evaluation and comparison with other methods. Best scores are in boldface.

Methods / Metrics	P@5	P@10	P@20	R-prec	MAP	bpref	MRR
Baseline models							
Balog's Model2 [1]	0.64	0.575	0.46	0.439	0.3915	0.3815	0.9167
BL	0.7	0.67	0.4975	0.4942	0.4803	0.4716	0.9
(vs **Balog's Model2**)	+8.57%	+14.18%	+7.54%	+11.17%	+18.48%*	+19.1%*	-1.85%
Enhanced models based on **BL**							
EnhancedBL [8]	0.73	0.68	0.535	0.5109	0.4906	0.4876	0.91
(vs **BL**)	+4.29%	+1.49%	+7.54%	+3.37%	+2.15%	+3.40%	+1.11%
Co-Ranking [30]	0.77	0.655	0.5125	0.4953	0.5049	0.4707	0.91
(vs **BL**)	+10%	-2.24%	+3.02%	+0.21%	+5.12%	-0.19%	+1.11%
JointHyp	**0.84**	**0.72**	**0.57**	**0.5513**	**0.5661**	**0.5394**	**1**
(vs **EnhancedBL**)	+15.07%*	+5.88%	+6.54%	+7.92%	+15.39%*	+10.63%*	+9.89%
(vs **Co-Ranking**)	+9.09%	+9.92%*	+11.22%*	+11.32%*	+12.13%*	+14.61%*	+9.89%

* indicates the improvement is statistically significant ($p < 0.05$).

Table 5: Experimental results of our proposed methods under various hypotheses. The percentages of relative improvements (%) over the baseline are also shown in the table. Best scores are in boldface.

Methods / Metrics	P@5	P@10	P@20	R-prec	MAP	bpref	MRR
BL	0.7	0.67	0.4975	0.4942	0.4803	0.4716	0.9
Individual hypothesis							
Hyp:Doc(G_D)	0.76	0.72	0.5525	0.5297	0.5318	0.5089	1
(vs **BL**)	+8.57%	+7.46%	+11.06%*	+7.18%*	+10.73%*	+7.93%*	+11.11%*
Hyp:Doc(G'_D)	0.73	0.695	0.5375	0.5238	0.5185	0.4997	0.95
(vs **BL**)	+4.29%	+3.73%	+8.04%*	+5.98%*	+7.96%*	+5.96%*	+5.56%
Hyp:Author	0.77	0.675	0.5525	0.5244	0.5302	0.5135	0.9017
(vs **BL**)	+10%*	+0.75%	+11.06%*	+8.27%*	+10.40%*	+8.9%*	+0.19%
Hyp:DocAuthor	0.73	0.685	0.52	0.4987	0.4921	0.4795	0.8833
(vs **BL**)	+4.29%	+2.24%	+4.52%*	+0.91%	+2.46%	+1.68%	-1.85%
Joint hypotheses							
JointHyp	**0.84**	**0.72**	**0.57**	**0.5513**	**0.5661**	**0.5394**	**1**
(vs **BL**)	+20%*	+7.46%	+14.57%*	+11.56%*	+17.87%*	14.39%*	+11.11%*
(vs **Hyp:Doc(G_D)**)	+10.53%*	0%	+3.17%	+4.08%	+6.45%*	+5.99%*	0%
(vs **Hyp:Author**)	+9.09%*	+6.67%	+3.17%	+5.14%	+6.76%*	+5.04%*	+10.91%*

* indicates the improvement is statistically significant ($p < 0.05$).

dependent community on top of **BL** model, which is quite relevant to our model.

Co-Ranking [30] can utilize the heterogeneous networks to rank authors and documents simultaneously, however, Co-Ranking cannot be directly used for expertise ranking since it is query-independent. Here we first constructed a query-specific subgraph based on the relevant documents and authors, and then performed Co-Ranking with empirically best parameters.

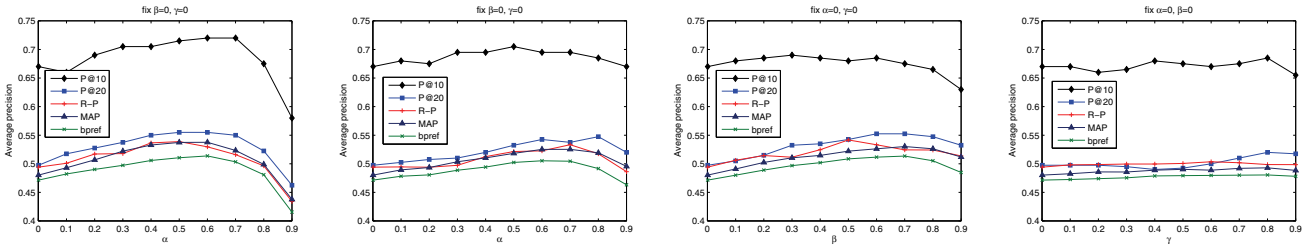
To make the comparison fair, we used the same benchmark dataset and reported the best performance of these models with *leave-one-out cross-validation*. Table 4 reports the evaluation results of these methods and our proposed model **JointHyp**. The relative improvement over baseline or other models is also shown in the table. As we can see, the results of our baseline model (BL) are much better than Balog's Model2, especially for MAP and bpref with 18%+ improvements. The reason is that BL model can benefit from the quality of documents in addition to the relevance which is used in Balog's Model2. This indicates our baseline model BL is a very competitive baseline. In terms of the enhanced models based on BL, both EnhancedBL and Co-Ranking can achieves slightly better performances over BL, since these two models integrate with some additional information. However, both enhanced models fail to outperform our proposed model JointHyp in all the metrics. Additionally, JointHyp gains significantly better performance over both EnhancedBL and Co-Ranking, e.g., 10%+ improvements for P@5, MAP and bpref. These observations demonstrate the effectiveness of our model by exploiting different information as well as heterogeneous networks successfully.

4.2.2 Comparison of Different Hypotheses

As discussed before, we develop three kinds of graph consistency hypotheses to enhance the baseline model by adding the regularization constraints. Based on the baseline model **BL**, we consider the following methods to validate three individual and the joint hypotheses, respectively.

- **Hyp:Doc(G_D)**: Exploiting document consistency hypothesis with citation graph G_D (Equation 3)
- **Hyp:Doc(G'_D)**: Exploiting document consistency hypothesis with co-conference graph G'_D (Equation 3)
- **Hyp:Author**: Exploiting co-authorship consistency hypothesis (Equation 5)
- **Hyp:DocAuthor**: Exploiting mutual document-author consistency hypothesis (Equation 9 & $\alpha = 0, \beta = 0$)
- **JointHyp**: Exploiting three hypotheses jointly (Equation 9).

For each case, we employ a grid-search and leave-one-out cross-validation approach [12] to learn the parameters



(a) Hyp:Doc(G_D) (b) Hyp:Doc(G'_D) (c) Hyp:Author (d) Hyp:DocAuthor

Figure 3: The effect of three individual hypotheses by varying α, β and γ.

that obtain the best performance. The experimental results of our proposed models are summarized in Table 5 where we show the precision and the relative improvement over baseline. For document consistency hypothesis, we consider two kinds of graphs, i.e., Hyp:Doc(G_D) and Hyp:Doc(G'_D), and we can see both of them have significant improvements over BL for most of the metrics, which indicates that document consistency is very useful. As expected, Hyp:Doc(G_D) appears to be more effective than Hyp:Doc(G'_D) since the directed citation graph provides more valuable information than the co-conference graph although the citation graph we used is not complete. We can imagine to achieve better performance with the complete citation graph. In the joint hypotheses testing, we only show the results using the citation graph due to space limit. For co-authorship consistency hypothesis, we observe that Hyp:Author also has significant improvement over BL, which shows the effectiveness of this hypothesis.

Now we test the mutual document-author consistency. Note that the propagation from documents' relevance to authors' expertise has been used in the baseline model, which is also the foundation of document-centric models for expert finding. The objective of this mutual hypothesis is to validate the effectiveness of the back propagation from authors to documents, i.e., whether the expertise scores of authors can be propagated to or reinforce the relevance scores of associated documents. As shown in Table 5, Hyp:DocAuthor only has slight improvement over BL. The possible reason is that the document-author consistency has partially used in BL, therefore additional value provided by the mutual document-author consistency may be limited in some sense. In general, the results of Hyp:Doc(G_D) and Hyp:Author are comparable, which are better than the mutual document-author consistency.

Since all the individual consistency hypotheses are effective, it is more interesting to check how the joint model JointHyp will perform by incorporating all the three hypotheses together. As we can see from Table 5, JointHyp outperforms BL, Hyp:Doc(G_D) and Hyp:Author with significant improvements[3]. This supports the joint hypothesis that different information and heterogeneous graphs should be modeled and exploited differently, so as to contribute additional value in the joint regularization model.

In summary, these individual consistency hypotheses make a great contribution to enhance the baseline probabilistic model. And meanwhile the joint regularization model can further improve the performance by incorporating all these hypotheses together.

[3]The improvement can be viewed as significant if the method has significant improvements based on MAP and bpref.

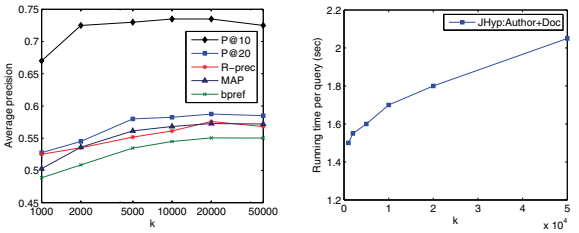
4.2.3 *Parameter Effect of Different Hypotheses*

In previous subsections, we learned the best parameters for the joint regularization framework using the grid-search and cross-validation approach. As for JointHyp, the best performance reported above is obtained by setting the parameters with $\alpha = 0.5$, $\beta = 0.6$ and $\gamma = 0.2$. Here we empirically investigate the parameter effect of different hypotheses. Validating the consistency of each of the three individual hypotheses is equivalent to explore one of the parameters α, β and γ after fixing other two parameters to 0, respectively. According to Table 1, if $\alpha = \beta = \gamma = 0$, the joint regularization framework reduces to the baseline model.

We first validate the *document consistency hypothesis* by varying α from 0 to 0.9 and fixing $\gamma = 0$, $\beta = 0$. Here the parameter α controls the balance between the document relevance consistency over the document graph (i.e., citation graph G_D or co-conference graph G'_D) and the initial relevance scores $\mathbf{x}^0$ of the probabilistic model. Figures 3(a) and (b) illustrate the experimental results of Hyp:Doc(G_D) and Hyp:Doc(G'_D) for different α, in which $\alpha = 0$ corresponds to the baseline **BL**. We can see the performance is improved over **BL** when incorporating the document relevance consistency ($\alpha > 0$). Generally, the performance becomes better with the increase of α until it puts too much weight on the term of document consistency. We observe that the performance is relatively stable and promising when α is set between 0.4 and 0.7. The first observation is that considering document consistency hypothesis could improve significantly over the baseline probabilistic model.

Then we evaluate the *co-authorship consistency hypothesis* by varying β from 0 to 0.9 and fixing $\alpha = 0$, $\gamma = 0$. The parameter β controls the importance of the co-authorship consistency. In Figure 3(c), the performance of Hyp:Author is improved along with the increase of β, and it is relatively stable between 0.4 and 0.7. These experiments indicate that the co-authorship consistency can be successfully used to refine expertise scores among authors.

Finally we evaluate the *mutual document-author consistency hypothesis* on top of **BL** by varying γ from 0 to 0.9, corresponding to Hyp:DocAuthor. The objective of this hypothesis is to validate whether the expertise scores of authors can be propagated to or reinforce the relevance scores of associated documents. As shown in Figures 3(d), the performance of Hyp:DocAuthor is stable and almost the same as the baseline. The possible reason is that experts usually have broad, diverse and cross-disciplinary research topics, for example, a data mining expert may have published many papers related to bioinformatics, therefore the propagation from experts to documents may result in topic drift. Another reason is that the document-author relationship has



(a) Performance (b) Running time per query

Figure 4: The performance and running time of model JHyp:Author+Doc by varying the parameter k with top-k relevant documents.

been implicitly used in the probabilistic model, therefore it may not contribute new information to the model and only small changes occur in the results by considering the propagation from authors to documents.

4.2.4 Detailed Results and Analysis

As mentioned in Section 3.3, we only need to retrieve a subset with top-k relevant documents, so as to construct a query-specific subgraph for efficient computation of our model. The parameter k used in previous experiments is set to 5,000. To investigate the effect of this parameter, we chose **JointHyp**, and evaluated it with 5 different k values from 1,000 to 50,000. The performance and running time[4] per query are depicted in Figures 4(a) and (b), respectively. We can see the performance becomes better for greater k, then levels off as $k > 10,000$. It is reasonable that more documents can better capture the complete expertise. In the meantime, the running time was found to increase less than linearly with the increase of k. Hence a good tradeoff is to set $k = 5,000$.

To gain a better insight into the proposed algorithm, we chose two queries "Probabilistic Relevance Model[5]" and "Kernel Methods[6]" as the example cases to show more detailed results. The top-10 author lists ranked by the models **BL** and **JointHyp** are shown in Table 6, where the relevant experts are in boldface. It is obvious that the results of JHyp:Author+Doc make more sense than the baseline. Taking the query "Kernel Methods" as an example, we find that JHyp:Author+Doc can boost some relevant researchers like "John Shawe-Taylor" and "Jason Weston" into top 10. Similarly, the results of another query shows the same observations. This is because of the document consistency and co-authorship consistency hypotheses. After looking into the details, one important observation is that the joint regularization framework can successfully constrain some irrelevant expertise and return mostly relevant results.

5. RELATED WORK

Generally, there are two principal models for expertise retrieval: profile-based model [16] and document-based model [1, 7, 10, 18]. In contrast, the document-based approach preserves all information contained in the collection and performs better than the profile-based approach. Besides these two categories, there are some other methods [13, 8, 23] proposed to extend the expertise retrieval. The organizational hierarchy [13] and community information [8] are utilized to enhance expert finding. However, these methods analyze the content of each document separately or merely extend the model with another additional kind of information. Our work is different from theirs, as we model and exploit heterogeneous networks in a joint regularization framework along with several hypotheses, which not only distinguishes the propagation between documents and authors, but also treats directed and undirected graphs in a different way.

Table 6: The top-10 experts retrieved by BL and JHyp:Author+Doc. Relevant experts are in boldface.

Baseline (**BL**)	**JointHyp**
Query: Kernel Methods	
Bernhard Schölkopf	**Bernhard Schölkopf**
Nello Cristianini	**Nello Cristianini**
Alexander J. Smola	**John Shawe-Taylor**
Francesco Camastra	**Alexander J. Smola**
Colin Campbell	Francesco Camastra
Tong Zhang	**Michael I. Jordan**
Stéphane Canu	**Colin Campbell**
Alessandro Verri	**Tong Zhang**
Shotaro Akaho	**Massimiliano Pontil**
Peter Sussner	**Jason Weston**
Query: Probabilistic Relevance Model	
Norbert Fuhr	**Norbert Fuhr**
Stephen E. Robertson	**Stephen E. Robertson**
Friedrich Gebhardt	**W. Bruce Croft**
M. E. Maron	**M. E. Maron**
J. L. Kuhns	**ChengXiang Zhai**
ChengXiang Zhai	**C. J. van Rijsbergen**
C. J. van Rijsbergen	Friedrich Gebhardt
Azadeh Shakery	**J. L. Kuhns**
W. Bruce Croft	**Chris Buckley**
Victor Lavrenko	**William S. Cooper**

The work is also concerned with predicting the quality of users or their generated content in social media by calculating their centrality in the organizational social network, including community-based question answering [2], document recommendation [31], collaborative tagging system [21, 11], and other online communities [5, 27]. Zhang et al. [27] analyzed an online forum, seeking to identify users with high expertise while ignoring the relevance of documents. The authors in [2] proposed a co-training idea that jointly models the quality of the author and the review. Noll et al. [21] proposed a graph-based method like HITS algorithm, through assigning different weights to link users and documents, for ranking users in a collaborative tagging system. However, most of these works are based on some generalization of PageRank [3] and HITS [14], and ignore the relevance of documents, which are inferior in performance to the state-of-the-art query-dependent expert finding methods.

There are some other related studies for combining heterogeneous information. PopRank model [20] is developed to integrate heterogenous relationships between objects, which showed authority score of one type of objects could be a combination of scores from different types of objects. Zhou et al. [30] proposed a method for co-ranking authors and their publications using several networks. In addition, Sun et al. [25] proposed a ranking-based clustering method for heterogeneous network analysis, which aims at clustering and ranking objects within clusters simultaneously. However, our work is different from theirs, as their tasks are mainly used in query-independent settings, while we focus on expertise ranking on query-dependent settings.

[4]The testing hardware environment is on a Windows PC with 2.4GHz CPU and 4GB physical memory.
[5]`http://dx.doi.org/10.1561/1500000019`
[6]`http://en.wikipedia.org/wiki/Kernel_methods`

Our joint regularization framework using graphs is closely related to graph-based semi-supervised learning [32, 29], which usually assumes label smoothness over the graph. These types of graph regularization methods have been successfully applied in topic modeling [19, 6], ad-hoc information retrieval [9], and review quality prediction [17] tasks. Mei et al. [19] extended the graph harmonic function for topic modeling with network regularization. Diaz [9] used score regularization to adjust ad-hoc retrieval scores from an initial retrieval. Although there have been some existing explorations, to our knowledge, the presented work is the first extensive study of a unified graph-based regularization method by explore the heterogeneous network in the field of expertise retrieval.

6. CONCLUSION AND FUTURE WORK

In this paper we studied the expertise ranking problem through modeling and exploiting heterogeneous network together with the textual content information. We formulate three types of hypotheses that capture different information in the heterogeneous network with respect to different types of edges. We not only mathematically model those hypotheses, but also validate them individually and jointly in a regularization framework. The experimental results show that our proposed approach can achieve significantly better results than the baseline and other state-of-the-art models.

The method we propose is quite generalizable and applicable for expertise ranking tasks in social media with heterogeneous network structure, for example, community-based question answering and online forum. In future work, it would be interesting to apply our techniques in these tasks.

7. ACKNOWLEDGMENTS

The work was supported in part by the U.S. National Science Foundation grants IIS-0905215, by the U.S. Army Research Laboratory under Cooperative Agreement No. W911NF-09-2-0053 (NS-CTA), and by two grants from the Research Grants Council of the Hong Kong SAR (No. CUHK 413210 and No. CUHK 415410). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

8. REFERENCES

[1] K. Balog, L. Azzopardi, and M. de Rijke. Formal models for expert finding in enterprise corpora. In *SIGIR*, pages 43–50, 2006.

[2] J. Bian, Y. Liu, D. Zhou, E. Agichtein, and H. Zha. Learning to recognize reliable users and content in social media with coupled mutual reinforcement. In *WWW*, pages 51–60, 2009.

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

[4] C. Buckley and E. M. Voorhees. Retrieval evaluation with incomplete information. In *SIGIR*, pages 25–32, 2004.

[5] C. S. Campbell, P. P. Maglio, A. Cozzi, and B. Dom. Expertise identification using email communications. In *CIKM*, pages 528–531, 2003.

[6] H. Deng, J. Han, B. Zhao, Y. Yu and C. X. Lin. Probabilistic topic models with biased propagation on heterogeneous information networks. In *KDD*, pages 1271-1279, 2011.

[7] H. Deng, I. King, and M. R. Lyu. Formal models for expert finding on dblp bibliography data. In *ICDM*, pages 163–172, 2008.

[8] H. Deng, I. King, and M. R. Lyu. Enhanced models for expertise retrieval using community-aware strategies. *IEEE Trans. SMC-B*, pages 93–106, 2012.

[9] F. Diaz. Regularizing ad hoc retrieval scores. In *CIKM*, pages 672–679, 2005.

[10] H. Fang and C. Zhai. Probabilistic models for expert finding. In *ECIR*, pages 418–430, 2007.

[11] 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.

[12] C.W. Hsu, C.C. Chang, and C.J. Lin. A practical guide to support vector classification. In *Technical Report*, 2003.

[13] M. Karimzadehgan, R. W. White, and M. Richardson. Enhancing expert finding using organizational hierarchies. In *ECIR*, pages 177–188, 2009.

[14] J. M. Kleinberg. Authoritative sources in a hyperlinked environment. *JACM*, 46(5):604–632, 1999.

[15] X. Liu, J. Bollen, M. L. Nelson, and H. V. de Sompel. Co-authorship networks in the digital library research community. *IPM*, 41(6):1462–1480, 2005.

[16] X. Liu, W. B. Croft, and M. B. Koll. Finding experts in community-based question-answering services. In *CIKM*, pages 315–316, 2005.

[17] Y. Lu, P. Tsaparas, A. Ntoulas, and L. Polanyi. Exploiting social context for review quality prediction. In *WWW*, pages 691–700, 2010.

[18] C. Macdonald and I. Ounis. Voting for candidates: adapting data fusion techniques for an expert search task. In *CIKM*, pages 387–396, 2006.

[19] Q. Mei, D. Cai, D. Zhang, and C. Zhai. Topic modeling with network regularization. In *WWW*, pages 101–110, 2008.

[20] Z. Nie, Y. Zhang, J.-R. Wen, and W.-Y. Ma. Object-level ranking: bringing order to web objects. In *WWW*, pages 567–574, 2005.

[21] M. G. Noll, C. man Au Yeung, N. Gibbins, C. Meinel, and N. Shadbolt. Telling experts from spammers: expertise ranking in folksonomies. In *SIGIR*, pages 612–619, 2009.

[22] J. M. Ponte and W. B. Croft. A language modeling approach to information retrieval. In *SIGIR*, 1998.

[23] E. Smirnova and K. Balog. A user-oriented model for expert finding. In *ECIR*, pages 580–592, 2011.

[24] I. Soboroff, A. de Vries, and N. Craswell. Overview of the TREC-2006 enterprise track. In *Proc. of TREC 2006*.

[25] Y. Sun, Y. Yu, and J. Han. Ranking-based clustering of heterogeneous information networks with star network schema. In *KDD*, pages 797–806, 2009.

[26] C. Zhai and J. D. Lafferty. A study of smoothing methods for language models applied to information retrieval. *ACM Trans. Inf. Syst.*, 22(2):179–214, 2004.

[27] J. Zhang, M. S. Ackerman, and L. Adamic. Expertise networks in online communities: structure and algorithms. In *WWW*, pages 221–230, 2007.

[28] J. Zhang, J. Tang, and J.-Z. Li. Expert finding in a social network. In *DASFAA*, pages 1066–1069, 2007.

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

[30] 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.

[31] D. Zhou, S. Zhu, K. Yu, X. Song, B. L. Tseng, H. Zha, and C. L. Giles. Learning multiple graphs for document recommendations. In *WWW*, pages 141–150, 2008.

[32] X. Zhu, Z. Ghahramani, and J. D. Lafferty. Semi-supervised learning using gaussian fields and harmonic functions. In *ICML*, pages 912–919, 2003.