DivRank: the Interplay of Prestige and Diversity in Information Networks

Qiaozhu Mei[1,2], Jian Guo[3], Dragomir Radev[1,2]
[1] School of Information, [2] Department of EECS, [3] Department of Statistics
University of Michigan
{qmei, guojian, radev}@umich.edu

ABSTRACT

Information networks are widely used to characterize the relationships between data items such as text documents. Many important retrieval and mining tasks rely on ranking the data items based on their centrality or prestige in the network. Beyond prestige, diversity has been recognized as a crucial objective in ranking, aiming at providing a non-redundant and high coverage piece of information in the top ranked results. Nevertheless, existing network-based ranking approaches either disregard the concern of diversity, or handle it with non-optimized heuristics, usually based on greedy vertex selection.

We propose a novel ranking algorithm, DivRank, based on a reinforced random walk in an information network. This model automatically balances the prestige and the diversity of the top ranked vertices in a principled way. DivRank not only has a clear optimization explanation, but also well connects to classical models in mathematics and network science. We evaluate DivRank using empirical experiments on three different networks as well as a text summarization task. DivRank outperforms existing network-based ranking methods in terms of enhancing diversity in prestige.

Categories and Subject Descriptors: H.2.8 [Database Applications]: Data Mining

General Terms: Algorithms

Keywords: Diversity, ranking, information networks, reinforced random walk

1. INTRODUCTION

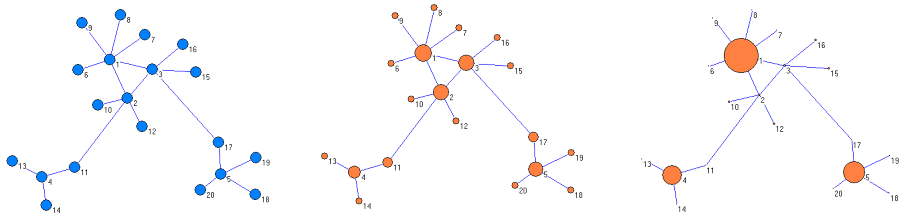
Consider the task of recommending three restaurants to a visitor. Without any prior information, a natural strategy is to recommend the three most famous ones, all of which happen to be seafood restaurants. However, the visitor could be a vegetarian, could prefer Chinese food, or could be allergic to seafood. A better strategy is thus to include something different in the recommendation, even though it is not as famous as the seafood restaurant it has replaced. A similar situation can be found in setting up the program committee of a conference, where an ideal committee should consist of prestigious researchers who cover all the related areas.

Many retrieval and mining tasks are concerned with finding the most important and/or relevant items from a large collection of data. Top ranked web pages are presented to the users of a search engine; top ranked job applicants are invited to on-site interviews; top ranked researchers are selected as the recipients of prestigious awards. Many ranking approaches have been proposed, ranging from pointwise weighting methods that use simple properties of each data item to network-based methods that utilize the relations among items, and to learning-to-rank methods that balance a lot of factors. Information networks, which characterize the relationships between the data items, have been playing an important role in these tasks. For instance, a search engine ranks web pages based on their prestige in a web hyperlink graph [14, 9]; researchers and scientific publications are ranked based on how well they are cited by other researchers. It is natural to assign a higher weight to data items that are referred to by many items, connected to many items, or on the paths between many items. These measures are known as centrality (or prestige) measures in general, with various instantiations like degree, closeness, betweenness [13], and more complicated measures such as the PageRank score [14] and the authority score [9]. These measures can be also combined with other features such as the relevance to a query.

However, the information need of a user usually goes beyond prestige or centrality. The diversity in top ranked results has been recognized as another crucial criteria in ranking. The top ranked items are expected to contain as little redundant information as possible, cover as many aspects as possible, or be as independent as possible. The need of diversity in ranking is even more urgent when the space of output is limited, for example in a mobile application.

Consider a toy example (illustrated in Figure 1(a)) which presents a network with 20 vertices. Vertex 1, 2, 3 and their neighbors are closely connected, while the ego-networks of vertex 4 and vertex 5 are loosely connected to the major community. Suppose the task is to find top-3 vertices to present the information of the whole network. If we rank the vertices using a prestige measure like degree or the PageRank (presented in Figure 1(b)), we can see that the top 3 ranked vertices are 1, 2, 3 respectively. All the three vertices are from the largest community and even form a clique by themselves. They are therefore likely to carry redundant information. Information of the two smaller communities centered at vertex 4 and vertex 5, however, does not present. A more desirable selection of the top-3 nodes should contain

KDD'10, July 25–28, 2010, Washington, DC, USA.
Copyright 2010 ACM 978-1-4503-0055-1/10/07 ...$10.00.



(a) An illustrative network. (b) Weighting with PageRank. (c) A diverse weighting.

Figure 1: An illustration of diverse ranking in a toy network.

diverse information, like in Figure 1(c). Vertex 1, 5, and 4 receive the majority weight, representing the three communities. Vertex 1, which represents the largest community, is ranked to the top. Although vertex 2 and 3 has a higher degree than vertex 5, they are ranked lower because vertex 1 has already partially covered their information.

A greedy vertex selection algorithm may achieve diversity by iteratively selecting the most prestigious vertex and then penalizing the vertices "covered" by the already selected ones. An example is the Maximum Marginal Relevance [3]. One may also consider first clustering the nodes (e.g., [17]) and then selecting centroids of clusters. However, it's difficult to predefine the number of clusters in this task. There lacks a principled objective and a unified process that automatically balances centrality and diversity.

In this paper, we propose a novel and unified process that balances prestige and diversity in ranking, based on a time-variant random walk in the network. The proposed model, called **DivRank** (abbreviation for *Diverse Rank*), introduces the *rich-gets-richer* mechanism to PageRank style random walks with reinforcements on transition probabilities. In contrast to the greedy vertex selection methods, DivRank provides a unified and intuitive stochastic process, as well as a principled optimization explanation. The process is well connected to a number of classical models in mathematics and network science, such as the *vertex-reinforced random walk*, the *Polya's Urn*, and the *preferential attachment*.

DivRank not only has a solid theoretical foundation, but also presents a strong empirical performance. The result presented in Figure 1(c) is actually generated using DivRank. We compare DivRank with a number of representative methods in literature using real world datasets and tasks. In all these tasks, DivRank outperforms the state-of-the-art methods in generating diverse top ranked results.

There are many potential applications of DivRank. The tasks presented in our experiments (i.e., ranking actors in social networks, ranking authors and publications, and text summarization) are by no means the only possible tasks. One may expect DivRank be applied in diversifying search results, snippet generation, keyword selection, mobile search, expert finding, and in various recommender systems.

The rest of the paper is organized as follows. In Section 2, we briefly introduce the task of ranking in information networks. In section 3, we formally introduce DivRank, including the general form and two practical approximations. We then provide an analytical discussion of DivRank in Section 4, followed by a comprehensive empirical analysis in Section 5. We discuss the related work in Section 6 and present our conclusions in Section 7.

2. RANKING VERTICES IN INFORMATION NETWORKS

In this section, we introduce the basic concepts and task of ranking vertices in networks, followed by the commonly used random walk processes for prestige measurement.

2.1 Information Networks

Let $G = (V, E)$ be a graph (or a network) where V is a finite set of vertices and E is a finite set of edges. We define an ordered pair (u, v) as an edge from vertex u to vertex v. When G is an *undirected* graph, we have $(u, v) = (v, u)$; when G is a *directed* graph, we have $(u, v) \neq (v, u)$. In a social network, V refers to a set of social actors (people) and E refers to the social ties between actors. In an information network, V and E broadly correspond to any type of information objects and the relationships between objects. We define the weight of an edge using $w(u, v)$. Note that when the edge corresponds to a citation between two documents or a hyperlink between two web pages, $w(u, v)$ takes a binary value. $w(u, v)$ could take any non-negative real value in other scenarios, e.g., when the edge corresponds to the similarity or cooccurrence of two objects, etc.

We then cast the task of ranking the vertices based on their *prestige* (or *centrality* in different contexts, which we will use interchangeably with prestige) in a network as finding a prestige function $f : V \to \mathbb{R}^+$. Beyond simple measures such as degree, recent research focuses on a family of centrality measures based on the stationary distribution of a random walk in the network, such as the well-known PageRank [14] and its counterpart in text networks, LexRank [5].

2.2 Prestige Ranking with Random Walks

A family of prestige measures in networks leverages the stationary distribution of a random walk in the network. A random walk defines a Markov chain in the given (either directed or undirected) network, where each vertex represents a state and a walk transits from one state to another based on a transition probability, denoted as $p(u, v)$. In other words, a random walk on G is defined by a transition probability function $p : V \times V \to [0, 1]$. Let us use $p_T(u)$ to denote the probability that the walk is at state u at time T. A standard random walk can be defined as

$$p_T(v) = \sum_{(u,v)\in E} p(u, v)p_{T-1}(u). \tag{1}$$

If the Markov chain is ergodic, $p_T(v)$ converges to a stationary distribution $\pi(v)$ which is commonly used to measure the importance of vertices.

Most existing random walk models assume that the transition probability $p(u,v)$ doesn't change over time, and instead can be estimated based on the topological structure of the network and/or the prior knowledge about the process. In a Web hyperlink graph, $p(u,v)$ can be estimated by

$$p(u,v) = \begin{cases} (1-d)\cdot\dfrac{1}{N} + d\cdot\dfrac{\mathbf{I}[(u,v)\in E]}{deg(u)}, & \text{if } deg(u) > 0 \\ \dfrac{1}{N}, & \text{o.w.,} \end{cases} \tag{2}$$

where d is a damping factor and $deg(u)$ is the out-degree of the web page u (the number of hyperlinks from u to other web pages). The Markov chain defined by $p(u,v)$ is ergodic. The stationary distribution of this random walk, $\pi(u)$, yields to the well known *PageRank* score for ranking web pages.

There are different ways to estimate $p(u,v)$. For example, in a general weighted graph, such as a document similarity graph, one can estimate $p(u,v)$ using $w(u,v)/\sum_{v\in V} w(u,v)$ to substitute $\mathbf{I}[(u,v)\in E]/deg(u)$[1] in Equation 2, where $w(u,v)$ is the weight of the edge (u,v). In another scenario where we have a prior distribution $p^*(v)$ $(s.t. \sum_v p^*(v) = 1)$, we can substitute $1/N$ in Equation 2 with $p^*(v)$. The stationary distribution of such a random walk then yields to *personalized PageRank*, or *topic-sensitive PageRank*.

In all these cases, we notice that the transition probabilities do not change throughout the random walk process. In other words, the corresponding Markov chain is time-homogenous. π assigns higher weights to vertices that are more prestigious. If one vertex is visited very frequently by the walk, all its neighbors are also more likely to be visited, thus inherit a prestige score from that vertex. This is known as a *regularization* process [23, 22], or a *smoothing* process [12] of scores in the network. In the scenario that vertices with high degrees are well connected, the top ranked vertices are likely to contain redundant information. In other words, the top ranked results is not diverse.

How to achieve diversity in a random walk? We may expect that there is not only a *smoothing* process between neighbors, but also a *competing* process. By doing this, we expect that rich nodes get richer over time and "absorb" the scores of its neighbors. In next section, we propose a principled way to facilitate this mechanism.

3. DIVRANK

We propose a new ranking algorithm in a network, called DivRank, which automatically balances centrality and diversity in the top ranked items. DivRank is motivated from a general time-variant random walk process known as the *vertex-reinforced random walk* in mathematics literature [15].

3.1 Vertex-Reinforced Random Walk

Time-homogenous random walks (e.g., PageRank) assume that the transition probabilities remain constant over time. In a real world random walk process, it is reasonable to consider the change of transition probabilities over time. Indeed, a visitor is more likely to walk to a museum that have already been visited by many visitors; people tend to join larger groups in a banquet; an actor accumulates prestige when acting in various movies, and the prestige in turn help her get even more opportunities. These can all be considered as random walk processes with time-variant transition probabilities. One particular family of time-variant random walk processes is known as the *vertex-reinforced random walks* (VRRW) in mathematics literature. The basic assumption is that the transition probability to one state from others is reinforced by the number of previous visits to that state.

Formally, let $p_0(u,v)$ be the transition probability prior to any reinforcement and let $N_T(v)$ be the number of times the walk has visited v up to time T. Then a VRRW can be defined sequentially as follows. First, we initialize $N_0(v) = 1$ for $v = 1,\dots,n$. Suppose we know the random walk stays at state u at time T, then at time $T+1$, the random walk moves to state v $(v = 1,\dots,n)$ with probability $p_T(u,v) \propto p_0(u,v)N_T(v)$ for any state u. In other words, $p_T(u,v)$ is reinforced by $N_T(v)$. The discussion of properties of vertex-reinforced random walks can be found in [15], which shows that, under some well-defined conditions, the score in VRRW converges to some stationary distribution almost surely.

3.2 The General Form of DivRank

We then introduce the general form of DivRank based on a similar reinforced random walk. Let $p_T(u,v)$ be the transition probability from any state u to any state v at time T. We can then define a family of time-variant random walk processes in which $p_T(u,v)$ satisfies

$$p_T(u,v) = (1-\lambda)\cdot p^*(v) + \lambda\cdot\frac{p_0(u,v)\cdot N_T(v)}{D_T(u)}, \tag{3}$$

where

$$D_T(u) = \sum_{v\in V} p_0(u,v)N_T(v). \tag{4}$$

Here, $p^*(v)$ is a distribution which represents the prior preference of visiting vertex v. When $p^*(v)$ is uniform, the left component is similar to the random jumping probabilities in PageRank. $p^*(v)$ could also be realized as a topic-sensitive distribution, similar to the personalized jumping probability in personalized PageRank. $p^*(v)$ could even be realized as the stationary distribution of a time-homogeneous random walk (e.g., PageRank). When $\lambda = 1$, Equation 3 yields to a standard vector-reinforced random walk.

$p_0(u,v)$ is the "organic" transition probability prior to any reinforcement, which can be estimated as in a regular time-homogenous random walk. After each step, the transition probabilities will be reinforced by the expected number of visits to each vertex. It is reasonable to assume that at any time, there is a probability that the walk stays at the current state, and this probability is reinforced by the number of visits at the current state. In other words, we assume there is always an "organic" link from a vertex to itself. We have

$$p_0(u,v) = \begin{cases} \alpha\cdot\dfrac{w(u,v)}{deg(u)}, & \text{if } u \neq v \\ 1-\alpha, & \text{if } u = v. \end{cases} \tag{5}$$

If the network is ergodic, after a sufficiently large T, the reinforced random walk defined by Equation 3 also converges to a stationary distribution $\pi(v)$. That is

$$\pi(v) = \sum_{v\in V} p_t(u,v)\pi(u), \qquad \forall t \geq T. \tag{6}$$

$\pi(v)$ is then used to rank the vertices in the information network, denoted as *DivRank*. Apparently, $\sum_{v\in V}\pi(v) = 1$.

[1] $\mathbf{I}(X)$ is an indicator function which returns 1 if the statement X is true and zero otherwise.

3.3 Efficient Approximations

In Section 3.2, we introduced the general form of DivRank based on a general reinforced random walk. Note that the expectation of $N_T(v)$ follows the recurrent formula

$$E[N_{T+1}(v)] = E[N_T(v)] + p_{T+1}(v) \ , \tag{7}$$

where $p_{T+1}(v){=}\sum_u p_T(u,v)p_T(u)$. It is easy to show that if $\pi(v)$ exists, we have $E[N_T(v)] \propto \pi(v)$ when T is sufficiently large. However, in DivRank $p_T(u,v)$ depends on $N_T(v)$ and tracking $N_T(v)$ is non-trivial. Efficient approximation is needed for practical applications.

In the original study of vertex-reinforced random walk, Pemantle proposed an approximation as follows [15]: Let $1 \ll L \ll T$, we can assume that the random walk process from time T to $T+L$ behaves as if $N_{T+L}(v)$ doesn't change over $N_T(v)$ since $L \ll T$. Therefore, the random walk in this period approximates a time-homogenous Markov chain with a fixed transition probability $p_T(u,v)$. Since $L \gg 1$, we may also assume that $N_{T+L}(v) - N_T(v)$ is proportional to the stationary distribution of such a Markov chain, $\pi_T(v)$. We can thus approximate $N_{T+L}(v)$ using $N_T(v) + L \cdot \pi_T(v)$.

This approximation, however, is still computationally inefficient. To find $\pi(v)$, one needs to compute the stationary distribution, $\pi_T(v)$, of many different Markovian random walks. In this section we propose two more practical approximations of DivRank.

Cumulative DivRank

One way to simplify the computation is to approximate $N_T(v)$ using $E[N_T(v)]$. In other words, we let $p_T(u,v) \propto p_0(u,v)E[N_T(v)]$. From Equation 7, we have

$$E[N_T(v)] \propto \sum_{t=0}^{T} p_t(v). \tag{8}$$

This is more efficient than the Pemantle approximation, since there is no need to compute the stationary distribution for every Markovian random walk. The underlining assumption is that the random walk approximately stays with every $p_t(v)$ for an equal period of time. We denote this approximation of DivRank as *cumulative DivRank.*

Pointwise DivRank

An even simpler approximation is to use $p_t(v)$ directly to approximate $E[N_t(v)]$. Indeed, when the random walk reaches the stationary status (at time T), $p_t(v)$ converges to $\pi(v)$. When the walk continues running for a sufficiently long time ($t \gg T$), $E[N_t(v)]$ is proportional to $\pi(v)$, or $p_t(v)$. With this simple approximation, we have

$$E[N_T(v)] \propto p_T(v). \tag{9}$$

We denote this simple approximation as *pointwise DivRank.* Equation 3 can then be simplified as

$$p_T(u,v) = (1-\lambda)\cdot p^*(v) + \lambda \cdot \frac{p_0(u,v)\cdot p_T(v)}{D_T(u)}, \tag{10}$$

where $D_T(u) = \sum_{v\in V} p_0(u,v)p_T(v)$. Below we use *DivRank* to refer to this *pointwise DivRank* unless specially noted.

4. ANALYTICAL DISCUSSION

We have now introduced the general form of DivRank and its two practical approximations. In this section, we present an analytical discussion of DivRank, including an optimization explanation and the connections to existing models.

4.1 The Optimization Explanation

What is the intuition behind DivRank, and what principle enhances diversity in ranking? Embedding Equation 3 into Equation 1, it is not hard to get

$$p_{T+1}(v) = (1-\lambda)p^*(v) + \lambda\sum_{u\in V}\frac{p_0(u,v)\cdot N_T(v)}{D_T(u)}p_T(u). \tag{11}$$

What does this process end up optimizing? Let us consider the scenario where the network is undirected (e.g., $w(u,v) = w(v,u)$) and $p_0(u,v)$ is estimated through $(1-\alpha)\cdot \mathbf{I}(u=v) + \alpha\cdot w(u,v)/deg(u)$. Let $D'_T(u) = \sum_{v\in V} w(u,v)N_T(u)N_T(v)$. Consider the following objective:

$$\begin{aligned} O_T(G) &= \lambda\cdot\sum_{u,v} w(u,v)N_T(u)N_T(v)(\frac{f_u}{D'_T(u)} - \frac{f_v}{D'_T(v)})^2 \\ &+ (1-\lambda)\sum_{v\in V}\frac{1}{D'_T(v)}(f_v - f^*_v)^2. \end{aligned} \tag{12}$$

By taking the partial derivative of f_v, it is not hard to show

$$\begin{aligned} \frac{\partial O_T(G)}{\partial f_v} &= 2\cdot\frac{1}{D'_T(v)}f_v - 2\cdot\frac{1}{D'_T(v)}[(1-\lambda)f^*_v \\ &+ \lambda\cdot\sum_u\frac{w(u,v)N_T(v)}{\sum_{v'}w(u,v')N_T(v')}\cdot f_u]. \end{aligned} \tag{13}$$

Let $f_v = p_{T+1}(v), f_u = p_T(u), f^*_v = p^*(v)$. It is easy to show that we can get Equation 11 from setting $\frac{\partial O_T(G)}{\partial f_v} = 0$.

This is to say, at every time T, the random walk defined in Equation 11 is attempting to improve the objective function in Equation 12. Let us analyze Equation 12. We can see that by minimizing the right component in the summation, f_v keeps close to the predefined value, f^*_v. By applying different f^*_v, one can incorporate various assumptions about f. For instance, one can assume that a reasonable f should rank nodes with a larger degree higher, and thus set f^*_v proportional to the degree of v.

The left component is more interesting. By minimizing the left component, the system regularizes the weights of vertices by dragging f_u/f_v towards $D'_T(u)/D'_T(v)$. In other words, this component in the objective function reflects the consistence of $f_v/D'_T(v)$ over the network, which is related to the consistency regularizer in machine learning literature [21, 22] but varies over time.

We know $D'_T(v) = \sum_{u\in V} w(u,v)N_T(u)N_T(v)$ and $N_0(u) = N_0(v)$. When the random walk starts, the optimization process is dragging f towards the degree distribution, which is exactly a Markovian random walk in the undirected graph. Nodes with a higher degree will get a higher weight, which in turn results in a larger accumulative N_T. When the random walk proceeds, the nodes which already have a high N_T tend to get an even higher weight. The self-link to a vertex guarantees that even if all the neighbors of v shrink, $D'_T(v)$ could still be large as long as $N_T(v)$ is large.

In other words, at time 0, the objective function in Equation 12 favors nodes with a higher centrality. As time goes by there emerges a **rich-gets-richer** phenomenon.

Indeed, the ratio of two adjacent nodes f_u/f_v is proportional to $D'_T(u)/D'_T(v)$ when the left component of Equation 12 is optimized. Originally, this ratio is proportional to the ratio of their degrees. As the random walk proceeds, this ratio tends to become more and more skewed. The self-links guarantee that adjacent nodes will "**compete**" for DivRank scores. Nodes already having high weights (thus higher $D'_T(v)$) are likely to "absorb" the weights of its neighbors directly, and the weights of neighbors' neighbors indirectly. As a result, the connectivity among the top ranked vertices tend to be low, thus enhances the diversity of information.

This provides an optimization explanation of DivRank on an undirected network. If the network is directed and there exists a reversible random walk ($\pi(u)p(u,v) = \pi(v)p(v,u)$), one could find a corresponding optimization explanation similar to the regularization framework on directed graphs in [22], but time-variant. However, it is generally hard to show an optimization explanation on an arbitrary directed graph, even for time-homogenous random walks like PageRank.

4.2 Connections to Other Models

Although DivRank is a novel model, it is well connected to a number of classical models in other contexts. First, when there is no jumping behavior (i.e., $\lambda = 1$), DivRank yields to a *vertex-reinforced random walk* with self-links.

If the graph is fully connected with uniform edge weight, the VRRW yields to a *Polya's Urn*. π converges to a Dirichlet compound multinomial distribution. Such a process has been used to model the word burstiness in text [11]. Indeed, the word sampling process in [11] can be viewed as a special case of a reinforced random walk on a network of words.

Furthermore, the process of DivRank is also related to *preferential attachment* models in network evolution, e.g., the Barabási-Albert model [2]. In a preferential attachment model, new nodes are more likely to attach to existing nodes that have already got a larger number of attachments, thus generates networks with a power-law degree distribution. Although DivRank does not model the evolution of networks (instead assumes that the topological structure is stable over time), we do see a related principle underneath, as well as a similar rich-gets-richer phenomenon. DivRank models the evolution of node weights and transition probabilities in the random walk in a "preferential transition" flavor. Indeed, one may call it a "preferential ranking" model.

There is another random walk based model in literature which improves the diversity in ranking. The *Grasshopper* model in [24] leverages an absorbing random walk. The model starts with a regular time-homogenous random walk. Every step the vertex with the largest weight is selected into the top ranked set, which is then set as an absorbing state. The model then reruns the random walk with absorbing states, and select the next vertex based on the expected number of visits to each node before absorption. In this way, *Grasshopper* also achieves diversity in the top ranked results. Comparing to DivRank, we can see Grasshopper takes a greedy approach of selecting vertices one by one. Such a process is similar to other greedy vertex selection methods like MMR, but with a "soft" penalization. In contrast, DivRank generates the ranking of vertices in a unified process, which balances prestige and diversity automatically. In the following section, we will compare DivRank with Grasshopper using real world datasets and tasks.

5. EXPERIMENTS

In this section, we evaluate the effectiveness of DivRank empirically. We select ranking tasks on three real world networks as well as a document summarization task.

5.1 Baselines and Evaluation Measures

A natural competitor of DivRank is *PageRank*, the random walk based centrality measure without diversity. We also compare with its two variations, namely the *personalized PageRank* and *LexRank* in the context of text summarization. We also compare DivRank with two diversity-aware methods, namely the Maximum Marginal Relevance (*MMR*) and *Grasshopper*. We believe that these baseline methods (except MMR) well represent the random walk based ranking methods in literature.

Note that MMR is originally proposed in a query-dependent context, thus it is not directly comparable to DivRank. In MMR, there is an explicit notion of relevance to the query. Although our setup is query-independent, we make a small variation to MMR so that it is comparable to PageRank, DivRank, and Grasshopper. The idea is to use PageRank score as the initial "relevance" score of MMR.

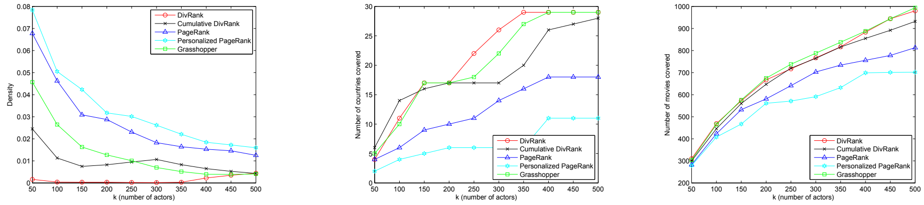
What is a reasonable general measurement of diversity? The recent "Redundancy, Diversity, and Interdependent Document Relevance" workshop at SIGIR 2009 concluded that "*there is no evaluation metric that seems to be universally accepted as the best for measuring the performance of algorithms that aim to obtain diverse rankings*" (quoted from [16]). Without a query, it is not feasible to apply the unified measures in retrieval that are related to "relevance" and "subtopics" [18, 4]. In our experiments, we thus present separate measures for prestige and diversity in the top-ranked results. [24] applies an ad hoc diversity measure in a particular context of ranking movie stars, i.e., the coverage of countries in the top ranked actors. Such a measure only assesses diversity indirectly and cannot be generalized to other datasets/tasks where such metadata is not available. In general, we need a measure that accounts for the redundancy of information in the top ranked vertices, even if we do not have any information other than the network itself.

In our experiments, we propose to leverage the *density* measure in network science. The density of a graph is defined as the number of edges (excluding self-links) presenting in the network divided by the maximal possible number of edges in the network. Formally, we can define

$$d(G) = \frac{\sum_{u \in V} \sum_{v \in V, u \neq v} \mathbf{I}[w(u,v) > 0]}{|V| \times (|V| - 1)}. \tag{14}$$

where $|V|$ is the number of vertices in G. Note that this general definition applies to both undirected graphs and directed graphs. Specifically, given the top-K ranked vertices, we can construct a subgraph of G, G_K, consisting of the top-K vertices and the edges among them. We then use the density of the subgraph, $d(G_K)$, as an *inverse* measure of diversity in top-K ranked vertices. Our assumption is that the *smaller* $d(G_K)$ is, the more independent the top-K vertices are, thus the less redundancy and *higher* diversity is contained by the top-ranked results (and vice versa).

For each task, we also define task-specific measure(s) to evaluate the quality (e.g., prestige) of the top-K vertices. Note that although it is difficult to develop a general measure that combines quality and diversity, in some tasks like document retrieval and text summarization, we can leverage the standard and unified measures in those contexts.



(a) Density (the lower the better). (b) Country coverage. (c) Movie coverage.

Figure 2: Comparison of network-based ranking methods in ranking IMDb stars. Parameters: λ (or d) = 0.9 in PageRank, Personalized PageRank, DivRank, Cumulative DivRank, and Grasshopper; $\alpha = 0.25$ in DivRank and Cumulative DivRank; .

5.2 Ranking in an Actor Social Network

Our first experiment considers a movie star social network extracted from the Internet Movie Database (IMDb[2]). It is exactly the same dataset used in Zhu et al. 2007 ([24]). The dataset covers 3452 unique actors/actresses from 1027 comedy movies produced between 2000 to 2006. We construct an actor social network by using all movie stars as vertices and weighting an edge between a pair of stars with the number of movies they co-starred in. Following [24], we also add a self-link to each actor with weight 1.

The task is to find the top-K actors that represent this social network. Ideally, the selected actors/actresses should be prestigious (appear at central positions in this network), and the top-K actors should also cover diverse groups of movie stars. Besides the density measure, we also include the “country coverage” measure which is used in [24]. To measure the “quality,” we follow [24] and compute the unique number of movies covered by the top-K stars (called “movie coverage”). The basic assumption is that the ideal top-K stars should cover as many unique movies as possible, and also cover comedians from different countries (which are assumed to represent different communities). Note that a prestigious actor is supposed to be starring in many movies, not necessarily co-starring with many actors. One could notice that the movie coverage measure is not a pure prestige measure, where there is also an implicit notion of diversity.

Following [24], we set the prior distribution $p^*(v)$ to be proportional to the number of movies that the actor has starred in. The same information is provided to Grasshopper and personalized PageRank.

We compare the performance of the random walk based ranking methods, including two approximations of DivRank, PageRank, personalized PageRank, and Grasshopper in Figure 2 using density, country coverage, and movie coverage. We set the jumping probability in all these methods to 0.1. Since there are separate measures for prestige and diversity, we didn’t tune the parameter α and simply set it 0.25. From Figure 2(a), we can clearly see that (pointwise) DivRank achieves the largest diversity in the top ranked results ($K <$ 500), followed by cumulative DivRank, then Grasshopper. When K is large enough (e.g., $K \sim 500$), these diversity generated by those three methods becomes similar. Density in top-K vertices generated by PageRank or personalized PageRank is clearly and consistently higher than those three methods, suggesting that PageRank-style random walks are not effective in enhancing diversity. Similar patterns can be observed in Figure 2(b), plotted with country coverage. The three diversity enhancing methods cover much more countries than PageRank and personalized PageRank, where cumulative DivRank provides the best coverage when K is smaller than 150, and pointwise DivRank provides the best coverage when K is between 200 and 400. One would ask whether the enhancement of diversity is at the expense of quality. From Figure 2(c), we see that the three diversity enhancing methods also generate higher movie coverage. The three methods perform comparably in terms of movie coverage, with Grasshopper slightly higher. All three methods present significantly larger movie coverage than PageRank and personalized PageRank. We also explored MMR, which generates similar results as Grasshopper [3]. It is interesting that personalized PageRank results in both the lowest diversity and the lowest movie coverage. This is because the movie coverage measure also partially accounts for diversity, and personalized PageRank works ineffectively with diversity.

Does the benefit of DivRank really come from the rich-gets-richer mechanism? To test this, we plot the score distribution of the top-K results by PageRank (left) and DivRank (right). From Figure 3, we can clearly see that the score distribution in top 100 results is much skewer in DivRank than in PageRank. Indeed, the riches (e.g., the top ranked vertices) have accumulated a significantly larger proportion of wealth (e.g., DivRank score [4]), which enhances the diversity.

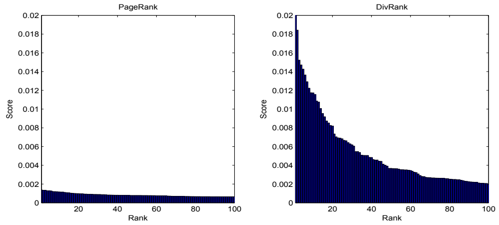


Figure 3: Score distribution in top 100 actors.

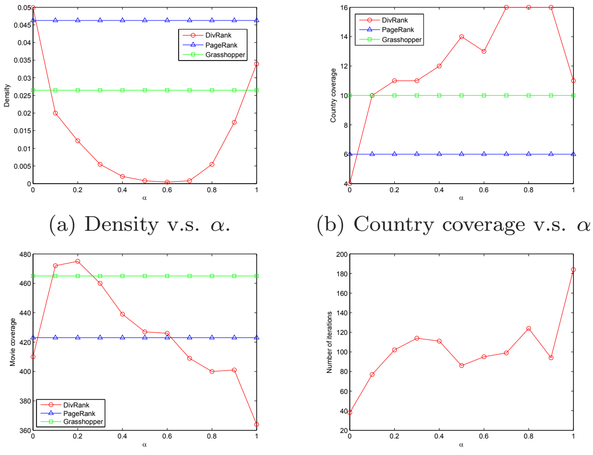
Although it is not meaningful to tune the parameters with multiple evaluation measures, we plot the performance of pointwise DivRank over different values of the parameter α, which controls the strength of self-links.

[2] http://www.imdb.com

[3] Note that the parameter setting of MMR is different from the random walk based methods. Without a unified measure, it is hard to set a “reasonable” parameter. Thus we don’t plot MMR.

[4] Note that the scores sum to 1.

Figure 4 plots the performance of pointwise DivRank using different values of α (K = 100). From Figure 4(a, b, c), we can see that density in top-K results is lower than PageRank and Grasshopper as long as the value of α is not extreme (i.e., close to 0 or 1). In the "comfort zone," a smaller α generates a higher movie coverage and a larger α generates a higher country coverage. The density measure is optimized when α is in the middle range of $(0, 1)$. In general, the performance is not sensitive when α in the middle range of $(0, 1)$. Figure 4(d) shows the number of iterations needed before DivRank converges. We can see that the convergence rate of DivRank is not sensitive to α. In general, the converging time presents the pattern that $PageRank < DivRank < Cumulated\ DivRank < Grasshopper$[5].



(a) Density v.s. α.

(b) Country coverage v.s. α.

(c) Movie coverage v.s. α

(d) Number of iterations

Figure 4: Effect of parameter α on the performance of DivRank (K = 100)

In summary, DivRank generates promising results in ranking movie stars, consistently outperforming PageRank and personalized PageRank in terms of diversity, country coverage, and movie coverage. It also provides top-K actors ($K < 500$) with comparable movie coverage but higher diversity, comparing to other diversity enhancing algorithms.

5.3 Ranking in Academic Networks

The results in ranking IMDb actors are promising. It is interesting to test whether the good performance of DivRank can be generalized to other datasets, especially to directed networks. To test this, we include a larger dataset with two directed networks extracted from an academic community.

The dataset is known as the ACL Anthology Network[6] (AAN), which contains an author citation network as well as a paper citation network constructed from 14912 papers collected in the ACL Anthology[7]. The author citation network covers 9641 authors, with each edge weighted by the number of times an author cited the work of the other. The paper citation network covers 11609 papers, with unit-weighted edges from each paper to the papers it cited.

These two networks are good representatives of directed networks. The task is to rank the most prestigious researchers (authors) and papers in the corresponding network, of course with diversity into consideration.

Like in the experiment with the actor social network, we use density to measure the *inverse* of diversity in the top-K results. To measure the quality, in this context the prestige of the top ranked authors, we leverage the well known *h-index* measure. By definition, if an author has published at most x papers which are cited for at least x times, the h-index of that author is x [8]. H-index provides a reasonable estimation of the author's impact in the community, which is widely used in ranking scholars in reality. We use the average h-index of the top-K authors as the prestige measure of ranking authors. For the paper citation network, we use the average number of citations of top-K papers as the prestige measure.

Please note that unlike movie coverage, both h-index and average citation purely measures prestige without any notion of diversity. For ranking papers in the paper citation network, we also include a measure that is similar to movie coverage, *impact coverage*, which counts the number of unique papers citing the top-K papers. This may be a better quality measure than the average number of citations.

The results are summarized in Figure 5. To simplify the illustration, we only plot three most representative methods: PageRank, DivRank, and Grasshopper. From Figure 5 (a) and (d), we can see that DivRank again effectively enhanced the diversity in top ranked results, which generates clearly and consistently sparser subgraph (i.e., higher diversity) with top-K items than PageRank and Grasshopper.

From Figure 5(b) and (e), we can see that in DivRank, the enhancement of diversity results in a lower average prestige when K is large, if the prestige measure considers every vertex independently. This is reasonable, since centrality-based methods (e.g., PageRank) always rank the most prestigious vertices to the top, which easily yields to the maximum independent prestige if the prestige measure correlates with the centrality. By enhancing diversity, one is optimizing the *marginal prestige* instead of the independent *independent prestige*. Like the movie coverage measure in ranking actors, the impact coverage in ranking papers (Figure 5(f)) confirms this intuition: both DivRank and Grasshopper produce a consistently higher impact coverage than PageRank, where DivRank is better for the top results ($K \leq 40$) and Grasshopper is better when $K > 40$. We've also tried MMR and it generates comparable results with Grasshopper.

It is worth mentioning that even with the independent prestige measure, the top-ranked results by DivRank outperform PageRank and Grasshopper. Indeed, when $K \leq 30$, we can clearly see from Figure 5 (b, e) that DivRank generates the highest average prestige among the three. This is desirable since for the user, the most useful (and preservable) information in a ranked list is always at the very top.

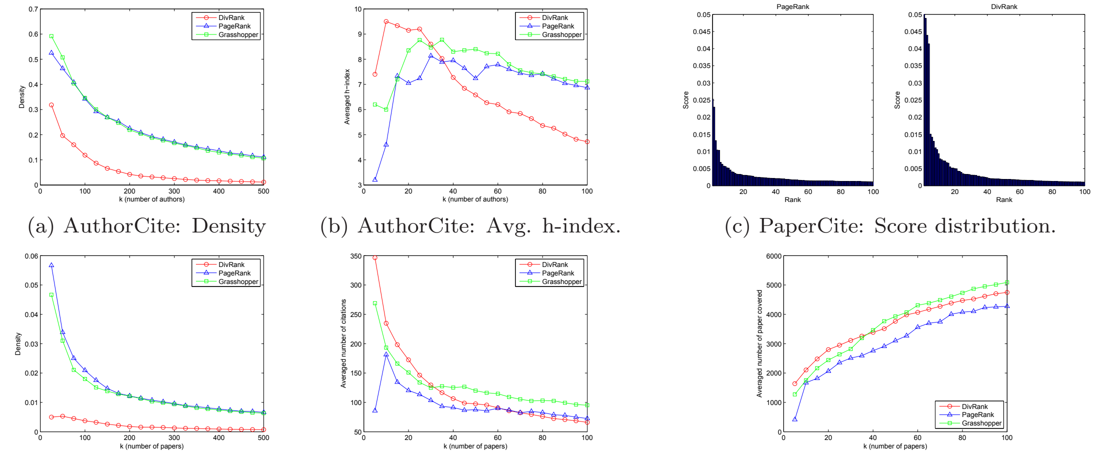
We also plot the score distribution of the top 100 papers in the paper citation network, presented in Figure 5(c). Again, we can see that the DivRank score distribution (in the right plot) presents a much skewer pattern than the PageRank score distribution (in the left plot) and the top ranked vertices have absorbed a much larger wealth. The pattern also appears in the author citation network.

This experiment proves that the effectiveness of DivRank generalizes to larger and directed networks.

[5] Note that Grasshopper selects one vertex at a step. Each step takes a few iterations to converge. The execution time is thus proportional to the size of output. Grasshopper is faster than DivRank if only the top few items are needed (e.g., $K < 10$), but slower when K is larger (e.g., $K \sim 100$).

[6] Downloadable at http://clair.si.umich.edu/clair/anthology/.

[7] http://www.aclweb.org/anthology-new/



(a) AuthorCite: Density

(b) AuthorCite: Avg. h-index.

(c) PaperCite: Score distribution.

(d) PaperCite: Density

(e) PaperCite: Avg. citations.

(f) PaperCite: Impact coverage.

Figure 5: Comparison of network-based ranking methods in AAN networks.
AuthorCite - Author Citation; PaperCite - Paper Citation. In (a), (d) (density comparison): the lower the better
Parameters: λ (or d) $= 0.9$ in PageRank, DivRank, and Grasshopper; $\alpha = 0.25$ in DivRank. $p^*(v)$ is uniform.

5.4 Text Summarization

In both the actor social network and the AAN academic networks, DivRank is evaluated using the task of prestige ranking directly. The good thing is that it gives us straightforward assessment of the performance of the ranking algorithms; the limitation, however, is that there isn't a unified metric that reasonably combines prestige and diversity.

This motivates us to evaluate DivRank by applying it to a task which has a well defined evaluation measure. Such a task should be able to be cast as a ranking problem, should be easy to evaluate, and should benefit from diversifying the top ranked results. We select document summarization as it does not depend on the input of queries. Document summarization is also chosen in [24] to evaluate Grasshopper.

We compare DivRank and its competitors using the Task 2 of the 2004 Document Understanding Conference (DUC)[8]. The goal is to generate a summary with no more than 100 words for a set of documents about each of the 50 included topics. We cast multi-document summarization as the problem of extracting the top-K sentences from the documents, and further as ranking the sentences based on a cosine similarity network. Because of the size limit of the summary, an ideal system should not only select the sentences most relevant to the topic, but also avoid redundant information in the selected sentences. We present each sentence in a topic with a TF-IDF vector space model. Then we create the graph of sentences by adding an unit-weighted edge between every two sentences if their cosine similarity is higher than 0.1. This is exactly the same setting in [5] and [24].

We evaluate the algorithms using ROUGE, the standard metric for text summarization [10]. ROUGE is a recall-based measure which compares the overlap between the system generated summary and the gold standard (human generated summary). Following [24], we also report ROUGE-1 score (unigram matching). The 50 topics were randomly split into a training set (with 30 topics) and a test set (with 20 topics). The former is used to select the optimal parameters in the models and the latter is used for evaluation.

Table 1: Results on DUC04 Task-2.

Method	Training		Testing	
	R-1	95% C.I.	R-1	95% C.I.
LR	0.359	[0.337, 0.381]	0.343	[0.318, 0.366]
PPR	0.378	[0.356, 0.398]	0.368	[0.350, 0.385]
MMR	0.363	[0.347, 0.379]	0.343	[0.318, 0.366]
GH	0.380	[0.360, 0.397]	0.356	[0.333, 0.378]
DR	**0.387**	[0.367, 0.404]	**0.379**	[0.366, 0.394]
CDR	0.384	[0.365, 0.401]	0.362	[0.342, 0.378]

R-1: Rouge-1. LR = LexRank (PageRank), PPR = personalized PageRank, MMR = marginal maximum relevance, GH = Grasshopper, DR = DivRank, CDR = cumulative DivRank. C.I.: Confidence Interval.

The averaged ROUGE-1 results on both training set and test set were listed in Table 1. Note that the MMR method uses LexRank [5], the counterpart of PageRank in summarization, as the "relevance" score. Consistent with [24], we provide the position information to personalized PageRank, Grasshopper, DivRank, and cumulative DivRank, by setting $p^*(v)$ according to the position of the sentence in a document. If the sentence v is the l^{th} sentence in a document, we set $p^*(v) \propto l^{-\beta}$, where β is a parameter.

From Table 1, we see that when tuning the parameters on the training data, DivRank performs the best, followed by cumulative DivRank and Grasshopper. Personalized PageRank performs reasonably well (better than LexRank), showing the importance of sentence position in document summarization. The good performance of DivRank generalizes well on the test data. The improvement over LexRank and Grasshopper is significant. This experiment shows that DivRank is effective when applied to text summarization.

6. RELATED WORK

The importance of enhancing diversity in ranking has been recognized in various contexts, including novelty detection [20], subtopic retrieval [18], diversifying search results [1, 6], recommender systems [25], and so on. [16] provides a summary of research problems and existing work on diversity in

[8] http://www-nlpir.nist.gov/projects/duc/guidelines/2004.html

information retrieval. Our paper studies the diversity problem in the context of ranking in information networks.

Centrality/Prestige ranking is a classic topic in network science. Many measures have been proposed, including degree, closeness, betweenness, impact domain, etc [13]. In the context of computer science, there have also been many well-accepted models and algorithms such as PageRank [14], HITS [9], LexRank [5], personalized PageRank [7], manifold regularization (e.g., [23]), etc. Most of these models are based on random walks or regularization on the network structure. However, none of them takes the diversity of information in the ranking into consideration.

The most related model to DivRank is Grasshopper, which is a vertex selection algorithm based on absorbing random walk [24]. Like MMR [3], Grasshopper takes a greedy approach to select one vertex at each step. Indeed, Grasshopper can be interpreted as a "soft" version of MMR. Instead of greedy vertex selection [3, 19, 24], DivRank generates the entire ranked list with one unified process, which automatically balances prestige and diversity based on a reinforced random walk model. To the best of our knowledge, DivRank is the first model which achieves this.

The theoretical framework of DivRank is related to quite a few classical models in mathematics literature, including the vertex-reinforced random walk [15] and the Polya's Urn. The rich-gets-richer mechanism of DivRank is also related to the preferential attachment (e.g., the Barabási-Albert model [2]) in network evolution which lays the foundation of scale-free networks, and word burstiness [11] in text mining.

7. CONCLUSION

We present DivRank, a novel ranking method in information networks that balances prestige and diversity. Unlike PageRank, DivRank employs a time-variant random walk process, which facilitates the rich-gets-richer mechanism in ranking. Diversity is achieved through the "competition" process between adjacent vertices. We show that DivRank has nice theoretical connections to a number of classical models in various contexts, such as vertex-reinforced random walk, the Polya's urn, and the preferential attachment. We also present a principled optimization explanation of DivRank.

Empirical experiments show that DivRank effectively enhances diversity in the top ranked results without the sacrifice of quality, which outperforms traditional greedy selection methods. The good performance generalizes well to directed graphs as well as real tasks that network-based ranking can be applied to.

There are many potential applications of DivRank. Given its good performance on paper citation networks, one can expect that DivRank facilitate ranking web pages in a web hyperlink graph. The results in text summarization also suggest potential applications in snippet generation and opinion extraction. One may imagine other applications such as keyword extraction and enhancing diversity in recommender systems.

An interesting future direction is to combine DivRank with other features in a learning-to-rank framework. Like PageRank, DivRank is proposed in a query-independent setup. It is interesting to extend DivRank to a query-dependent scenario, which leads to applications like subtopic retrieval, search result diversification, and expert finding.

Acknowledgements

We thank Prof. Xiaojin Zhu and Andrew Goldberg for sharing the code and data of Grasshopper. We thank Dr. Dengyong Zhou for his useful comments.

8. REFERENCES

[1] R. Agrawal, S. Gollapudi, A. Halverson, and S. Ieong. Diversifying search results. In *WSDM '09: Proceedings of the Second ACM International Conference on Web Search and Data Mining*, pages 5–14, 2009.

[2] A.-L. Barabasi and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.

[3] J. Carbonell and J. Goldstein. The use of mmr, diversity-based reranking for reordering documents and producing summaries. In *SIGIR '98: Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 335–336, 1998.

[4] C. L. Clarke, M. Kolla, G. V. Cormack, O. Vechtomova, A. Ashkan, S. Büttcher, and I. MacKinnon. Novelty and diversity in information retrieval evaluation. In *SIGIR '08: Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 659–666, 2008.

[5] G. Erkan and D. R. Radev. Lexrank: graph-based lexical centrality as salience in text summarization. *J. Artif. Int. Res.*, 22(1):457–479, 2004.

[6] S. Gollapudi and A. Sharma. An axiomatic approach for result diversification. In *WWW '09: Proceedings of the 18th international conference on World wide web*, pages 381–390, 2009.

[7] T. H. Haveliwala. Topic-sensitive pagerank. In *WWW '02: Proceedings of the 11th international conference on World wide web*, pages 517–526, 2002.

[8] J. Hirsch. An index to quantify an individual's scientific research output. *PNAS*, 102(46):16569–16572, 2005.

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

[10] C.-Y. Lin and E. Hovy. Automatic evaluation of summaries using n-gram co-occurrence statistics. In *NAACL '03: Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology*, pages 71–78, 2003.

[11] R. E. Madsen, D. Kauchak, and C. Elkan. Modeling word burstiness using the dirichlet distribution. In *ICML '05: Proceedings of the 22th international conference on Machine learning*, pages 545–552, 2005.

[12] Q. Mei, D. Zhang, and C. Zhai. A general optimization framework for smoothing language models on graph structures. In *SIGIR '08: Proceedings of the 31th annual international ACM SIGIR conference on Research and development in informaion retrieval*, pages 611–618, 2008.

[13] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45:167–256, 2003.

[14] L. Page, S. Brin, RajeevMotwani, and TerryWinograd. The pagerank citation ranking: Bringing order to the

web. *Technical report, Stanford Digital Library Technologies Project*, 1998.

[15] R. Pemantle. Vertex reinforced random walk. *Prob. Th. and Rel. Fields*, pages 117–136, 1992.

[16] F. Radlinski, P. N. Bennett, B. Carterette, and T. Joachims. Redundancy, diversity and interdependent document relevance. *SIGIR Forum*, 43(2):46–52, 2009.

[17] J. Shi and J. Malik. Normalized cuts and image segmentation. In *CVPR '97: Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition*, pages 731–737, 1997.

[18] C. X. Zhai, W. W. Cohen, and J. Lafferty. Beyond independent relevance: methods and evaluation metrics for subtopic retrieval. In *SIGIR '03: Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, pages 10–17, 2003.

[19] B. Zhang, H. Li, Y. Liu, L. Ji, W. Xi, W. Fan, Z. Chen, and W.-Y. Ma. Improving web search results using affinity graph. In *SIGIR '05: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in informaion retrieval*, pages 504–511, 2005.

[20] Y. Zhang, J. Callan, and T. Minka. Novelty and redundancy detection in adaptive filtering. In *SIGIR '02: Proceedings of the 25th annual international ACM SIGIR conference on Research and development in informaion retrieval*, pages 81–88, 2002.

[21] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Schölkopf. Learning with local and global consistency. In *NIPS '04*. 2004.

[22] D. Zhou, J. Huang, and B. Schölkopf. Learning from labeled and unlabeled data on a directed graph. In *Proceedings of the 22th international conference on Machine learning*, pages 1036–1043, 2005.

[23] D. Zhou, J. Weston, A. Gretton, O. Bousquet, and B. Schölkopf. Ranking on data manifolds. In *NIPS '04*. 2004.

[24] X. Zhu, A. Goldberg, J. Van Gael, and D. Andrzejewski. Improving diversity in ranking using absorbing random walks. In *NAACL-HLT 2007*, pages 97–104, April 2007.

[25] C.-N. Ziegler, S. M. McNee, J. A. Konstan, and G. Lausen. Improving recommendation lists through topic diversification. In *WWW '05: Proceedings of the 14th international conference on World wide web*, pages 22–32, 2005.