Diversifying Citation Recommendations

ONUR KÜÇÜKTUNÇ*, ERIK SAULE†, KAMER KAYA* and ÜMİT V. ÇATALYÜREK*,
* The Ohio State University
† University of North Carolina at Charlotte

Literature search is one of the most important steps of the academic research. With more than one hundred thousand papers published each year just in computer science, performing a complete literature search becomes a herculean task. Some of the existing approaches and tools for literature search cannot compete with the characteristics of today's literature, and they suffer from ambiguity and homonymy. Techniques based on the citation information are more robust to the mentioned issues. Following the idea, we have recently built a web service called the**advisor** which provides personalized recommendations to the researchers based on their papers-of-interest. Since most recommendation methods may return redundant results, diversifying the results of the search process is necessary to increase the amount of information one can reach via an automized search. This paper targets the problem of result diversification in citation-based bibliographic search, assuming that the citation graph itself is the only information available, and no categories or intents are known. The contribution of this work is three-fold: We survey various random-walk-based diversification methods and enhance them with the *direction awareness* property to allow the users to reach either old, foundational (possibly well-cited and well-known) research papers or recent (most likely less-known) ones. Next, we propose a set of novel algorithms based on vertex selection and query refinement. A set of experiments with various evaluation criteria shows that the proposed γ-RLM algorithm performs better than the existing approaches and is suitable for real-time bibliographic search in practice.

Categories and Subject Descriptors: H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval

General Terms: Algorithms, Experimentation

Additional Key Words and Phrases: Bibliographic search, diversity, direction awareness

ACM Reference Format:
Onur Küçüktunç, Erik Saule, Kamer Kaya, and Ümit V. Çatalyürek. 2013. Diversifying Citation Recommendations. *ACM Trans. Intell. Syst. Technol.* 0, 0, Article 00 (July 2013), 21 pages.
DOI:http://dx.doi.org/10.1145/0000001.0000001

This work was partially supported by the NHI/NCI grant R01CA141090 and NSF grant CNS-0643969.
This work was done while E. Saule was at the Ohio State University.
Author's addresses: O. Küçüktunç, Department of Computer Science and Engineering, The Ohio State University; email: kucuktunc.1@osu.edu; E. Saule, Department of Computer Science, University of North Carolina at Charlotte; email: esaule@uncc.edu; and K. Kaya and Ü. V. Çatalyürek, Department of Biomedical Informatics, The Ohio State University; emails: {kamer, umit}@bmi.osu.edu.

DOI:http://dx.doi.org/10.1145/0000001.0000001

1. INTRODUCTION

The academic community has published millions of research papers to date, and the number of new papers has been increasing with time. With more than one hundred thousand papers published in computer science each year, performing a complete literature search becomes a herculean task. Developing tools that help researchers to find relevant papers has been of interest for the last thirty years.

Some of the existing approaches and tools for literature search cannot compete with the characteristics of today's literature. For example, keyword-based approaches suffer from the confusion induced by different names of identical concepts in different fields. (For instance, *partially ordered set* or *poset* are also often called *directed acyclic graph* or *DAG*). Conversely, two different concepts may have the same name in different fields (for instance, *hybrid* is commonly used to specify software hybridization, hardware hybridization, or algorithmic hybridization). These two problems may drastically increase the number of suggested but unrelated papers.

Bibliographic search techniques based only on the citation information do not suffer from the above-mentioned problems [Gori and Pucci 2006; Kessler 1963; Lao and Cohen 2010; Lawrence et al. 1999; Li and Willett 2009; Liang et al. 2011; Ma et al. 2008; Small 1973] since they do not use textual information. Furthermore, it has been shown that most of the relevant information is contained within the citation graph [Strohman et al. 2007], and there is already a correlation between citation similarities and text similarities of the papers [Peters et al. 1995; Salton 1963].

Following the idea of using citation similarities for bibliographic search, we have recently built a web service called the**advisor**[1] [Kucuktunc et al. 2012a; Kucuktunc et al. 2012b]. It takes a bibliography file containing a set of papers, i.e., *seeds*, as an input to initiate the search. The algorithms employed by the**advisor** have the *direction-awareness* functionality which allows the user to specify her interest in classical or recent papers. Taking this criteria into account, the service returns a set of suggested papers ordered with respect to a ranking function which is a variant of personalized PageRank. After obtaining the results, the user can give relevance feedback to the system, and if desired, the output set is refined.

Diversifying the results of the search process is an important task to increase the amount of information one can reach via an automized search tool. There exists many recommender systems that personalize the output with respect to user's query/history. For several applications *personalization* can be an important limitation while reaching all the relevant information [Drosou and Pitoura 2010], and *diversification* can be used to increase the coverage of the results and hence, improve user satisfaction [Agrawal et al. 2009; Clarke et al. 2008; Gollapudi and Sharma 2009; Mei et al. 2010].

Most diversification studies in the literature rely on various assumptions, e.g., items and/or queries are categorized beforehand [Welch et al. 2011], or there is a known distribution that specifies the probability of a given query belonging to some categories [Agrawal et al. 2009]. In the context of information retrieval or web search, since the search queries are often *ambiguous* or *multifaceted*, a query should represent the *intent* of an average user with a probability distribution [Welch et al. 2011]. Intent-aware methods in the literature aim to cover various relevant categories with one or more objects.

In this work, we target the bibliographic search problem assuming that the citation graph itself is the only information we have, and no categories or intents are available. Hence, we aim to diversify the results of the citation/paper recommendation process with the following objectives in mind: (1) the direction awareness property is kept,

[1] http://theadvisor.osu.edu/

(2) the method should be efficient enough to be computable in real time, and (3) the results are relevant to the query and also diverse among each other. The contribution of this work is three-fold:

— We survey various random-walk-based diversification methods (i.e., GrassHopper [Zhu et al. 2007], DivRank [Mei et al. 2010], and Dragon [Tong et al. 2011]) and enhance them with the direction awareness property.
— We propose new algorithms based on vertex selection (LM, γ-RLM) and query refinement (GSPARSE).
— We perform a set of experiments with various evaluation criteria including relevance metrics, diversity metrics and intent-aware metrics. The experiments show that the proposed γ-RLM algorithm is suitable in practice for real-time bibliographic search.

All of the algorithms in this paper are implemented and tested within the**advisor** and the best one (γ-RLM) is currently being used to power the system.

2. BACKGROUND

2.1. Related Work

Graph-based Citation Recommendation. Paper recommendation based on citation analysis has been a popular problem since the '60s. There are methods that only take local neighbors (i.e., citations and references) into account, e.g., bibliographic coupling [Kessler 1963], cocitation [Small 1973], and CCIDF [Lawrence et al. 1999]. Recent studies, however, employ graph-based algorithms, such as Katz [Liben-Nowell and Kleinberg 2007], random walk with restart [Tong et al. 2006], or well-known PageRank (PR) algorithm [Brin and Page 1998] to investigate the whole citation network. PaperRank [Gori and Pucci 2006], ArticleRank [Li and Willett 2009], and Katz distance-based methods [Strohman et al. 2007] are typical examples.

Ranking with Personalized PageRank (PPR) is a good way of finding the probability of the papers' relevance for a given query. However, these algorithms treat the citations and references in the same way. This may not lead the researcher to recent and relevant papers if she is more interested in those. Old and well cited papers have an advantage with respect to the relevance scores since they usually have more edges in the graph. Hence the graph tends to have more and shorter paths from the seed papers to old papers. We previously defined the class of *direction aware* algorithms based on personalized PageRank, which can be tuned to reach a variety of citation patterns, allowing them to match the patterns of recent or traditional documents [Kucuktunc et al. 2012b]. We give the details of PageRank-based algorithms in Section 2.3.

Result Diversification on Graphs. The importance of diversity in ranking has been discussed in various data mining fields, including text retrieval [Carbonell and Goldstein 1998], recommender systems [Ziegler et al. 2005], online shopping [Vee et al. 2008], and web search [Clarke et al. 2008]. The topic is often addressed as a multi-objective optimization problem [Drosou and Pitoura 2010], which is shown to be NP-hard [Carterette 2009], and, therefore, some greedy [Agrawal et al. 2009; Zuccon et al. 2012] and clustering-based [Liu and Jagadish 2009] heuristics were proposed. Although there is no single definition of diversity, different objective functions and axioms expected to be satisfied by a diversification system were discussed in [Gollapudi and Sharma 2009].

Diversification of the results of random-walk-based methods on graphs only attracted attention recently. GRASSHOPPER is one of the earlier algorithms and addresses diversified ranking on graphs by vertex selection with absorbing random walks [Zhu et al. 2007]. It greedily selects the highest ranked vertex at each step and

turns it into a sink for the next steps. Since the algorithm has a high time complexity, it is not scalable to large graphs. DIVRANK [Mei et al. 2010], on the other hand, combines the greedy vertex selection process in one unified step with the vertex reinforced random walk model. This algorithm updates the transition matrix at each iteration with respect to the current or cumulative ranks of the nodes to introduce a *rich-gets-richer* mechanism to the ranking. But since the method updates the full transition matrix at each iteration, more iterations are needed for convergence; therefore, the computation cost increases. The shortcomings of those techniques were discussed in [Li and Yu 2011] in detail. [Tong et al. 2011] formalizes the problem from an optimization viewpoint, proposes the *goodness* measure to combine relevancy and diversity, and presents a near-optimal algorithm called DRAGON. These algorithms are further discussed in Section 3.

Coverage-based methods (such as [Kucuktunc et al. 2013; Li and Yu 2011]) are also interesting for diversification purposes; however, they do not preserve the direction awareness property of the ranking function. Since our aim is to diversify the results of our paper recommendation service, we omitted the results of those coverage-based methods in our experiments.

2.2. Problem Formulation

Let $G = (V, E)$ be a directed citation graph where $V = \{v_1, \ldots, v_n\}$ is the vertex set and E, the edge set, contains an edge (u, v) if paper u references paper v. We say that v is a reference of u, and that u is a citation of v. Let $\delta^+(u)$ and $\delta^-(u)$ be the number of references of and citations to paper u, respectively. We define the weight of an edge, $w(u, v)$, based on how important the citation is; however, for the sake of simplicity we take $w(u, v) = 1$ for all $(u, v) \in E$. Therefore, the nonsymmetric matrix $\mathbf{W} : V \times V$ becomes a 0-1 matrix. Table I summarizes the notation used throughout the paper.

We target the problem of paper recommendation assuming that the researcher has already collected a list of papers of interest [Kucuktunc et al. 2012b]. Therefore, the objective is to return papers that extend that list: given a set of m seed papers $\mathcal{Q} = \{q_1, \ldots, q_m\}$, s.t. $\mathcal{Q} \subseteq V$, and a parameter k, return top-k papers which are relevant to the ones in $\mathcal{Q}$. With the diversity objective in mind, we want to recommend papers to be not only relevant to the query set $\mathcal{Q}$, but also covering different topics around the query set.

2.3. PageRank, Personalized PageRank, and Direction-aware Personalized PageRank

Let $G' = (V, E')$ be an undirected graph of the citation graph, $p(u, v)$ be the transition probability between two nodes (states), and d be the damping factor.

PageRank (PR) [Brin and Page 1998]: We can define a random walk on G' arising from following the edges (links) with equal probability and a random restart at an arbitrary vertex with $(1-d)$ teleportation probability. The probability distribution over the states follows the discrete time evolution equation

$$\mathbf{p}_{t+1} = \mathbf{P}\ \mathbf{p}_t, \tag{1}$$

where $\mathbf{p}_t$ is the vector of probabilities of being on a certain state at iteration t, and $\mathbf{P}$ is the transition matrix defined as:

$$\mathbf{P}(u, v) = \begin{cases} (1-d)\frac{1}{n} + d\frac{1}{\delta(v)}, & \text{if } (u, v) \in E' \\ (1-d)\frac{1}{n}, & \text{otherwise.} \end{cases} \tag{2}$$

Table I.
Notation

	Symbol	Definition
Graph	G	directed citation graph, $G=(V,E)$
	G'	undirected citation graph, $G'=(V,E')$
	n	$\|V\|$, number of vertices
	$w(u,v)$	weight of the edge from u to v
	$\mathbf{W}$	weight matrix
	$\delta^-, \delta^+(v)$	# incoming or outgoing edges of v
	$\delta(v)$	$\delta^-(v)+\delta^+(v)$, # neighbors of v
	$d(u,v)$	shortest distance between u and v in G'
	$N_\ell(S)$	ℓ-step expansion set of $S \subseteq V$
Query	$\mathcal{Q}$	a set of seed papers $\{q_1,\ldots,q_m\}$, $\mathcal{Q} \subseteq V$
	m	$\|\mathcal{Q}\|$, number of seed papers
	k	required number of results, $k \leq n$
	R	a set of recommended vertices, $R \subseteq V$
	d	damping factor of RWR, $0<d\leq 1$
	κ	direction-awareness parameter, $0 \leq \kappa \leq 1$
	γ	relaxation parameter of γ-RLM
Random walk	p^*	prior probability distribution
	t	iteration, or timestamp
	$\mathbf{p}_t$	probability vector at iteration t
	η_t	vector of number of visits at iteration t
	$\mathbf{A}$	symm. $n \times n$ transition matrix based on G
	$\mathbf{A}'$	struct.-symm. $n \times n$ trans. matrix based on G'
	$\mathbf{P}$	$n \times n$ transition matrix
	π	$\mathbf{p}_\infty$, stationary probability vector, $\sum \pi(.)=1$
	ϵ	convergence threshold
Measures	S	a subset of vertices, $S \subseteq V$
	$\hat{S}$	top-k results according to π
	$rel(S)$	normalized relevance of the set
	$diff(S)$	difference ratio of two sets
	$use(S)$	usefulness of the set
	$dens_\ell(S)$	ℓ-step graph density
	$\sigma_\ell(S)$	ℓ-expansion ratio

If the network is ergodic (i.e., irreducible and non-periodic), (1) converges to a stationary distribution $\pi = \mathbf{P}\pi$ after a number of iterations. And the final distribution π gives the PageRank scores of the nodes based on *centrality*.

In practice, the algorithm is said to be converged when the probability of the papers are stable. Let

$$\Delta_t = (\mathbf{p}_t(1) - \mathbf{p}_{t-1}(1), \ldots, \mathbf{p}_t(n) - \mathbf{p}_{t-1}(n)) \tag{3}$$

be the difference between probability distributions at iteration t and $t-1$. The process is in the *steady state* when the L2 norm of Δ_t is smaller than the convergence threshold ϵ.

Personalized PageRank (PPR) [Haveliwala 2002]: In our problem, a set of nodes $\mathcal{Q}$ was given as a query, and we want the random walks to teleport to only those given nodes. Let us define a prior distribution p^* such that:

$$p^*(u) = \begin{cases} 1/m, & \text{if } u \in \mathcal{Q} \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

If we substitute the two $(1/n)$s in (2) with p^*, we get a variant of PageRank, which is known as *personalized PageRank* or *topic-sensitive PageRank* [Haveliwala 2002]. PPR scores can be used as the relevance scores of the items in the graph. The rank of each seed node is reset after the system reaches to a steady state, i.e., $\forall q \in \mathcal{Q},\ \pi_q \leftarrow 0$, since

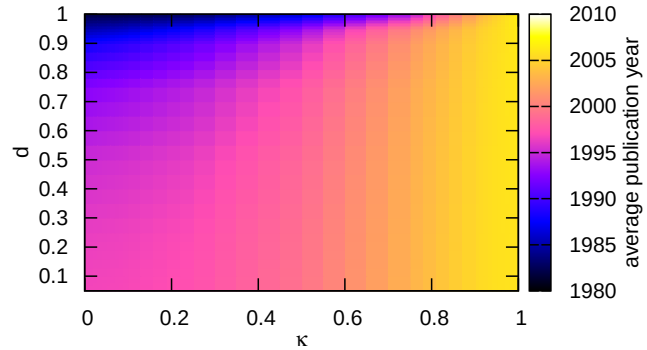


Fig. 1. Average publication year of top-10 recommendations by DARWR based on d and κ.

the objective is to extend $\mathcal{Q}$ with the results.

Direction-aware RWR (DARWR) [Kucuktunc et al. 2012b]: We defined a *direction awareness* parameter $\kappa \in [0, 1]$ to obtain more recent or traditional results in the top-k documents [Kucuktunc et al. 2012b]. Given a query with inputs k, a *seed* paper set $\mathcal{Q}$, damping factor d, and direction awareness parameter κ, Direction-aware Random Walk with Restart (DARWR) computes the steady-state probability vector π. The ranks of papers after iteration t is computed with the following linear equation:

$$\mathbf{p}_{t+1} = p^* + \mathbf{A}\mathbf{p}_t, \tag{5}$$

where p^* is an $n \times 1$ restart probability vector calculated with (4), and $\mathbf{A}$ is a structurally-symmetric $n \times n$ matrix of edge weights, such that

$$a_{ij} = \begin{cases} \frac{d(1-\kappa)}{\delta^+(i)}, & \text{if } (i,j) \in E \\ \frac{d\kappa}{\delta^-(i)}, & \text{if } (j,i) \in E \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

The transition matrix $\mathbf{P}$ of the RWR-based methods is built using $\mathbf{A}$ and p^*; however, the edge weights in rows can be stored and read more efficiently with $\mathbf{A}$ in practice [Kucuktunc et al. 2012a].

Figure 1 shows that the direction-awareness parameter κ can be adjusted to reach papers from different years with a range from late 1980's to 2010 for almost all values of d. In our service, the parameter κ can be set to a value of user's preference. It allows the user to obtain recent papers by setting κ close to 1, or older papers by setting κ close to 0.

3. DIVERSIFICATION METHODS

We classify the diversification methods for the paper recommendation problem based on whether the algorithm needs to rank the papers only once or multiple times. The first set of algorithms run a ranking function (e.g., PPR, DARWR, etc.) once and select a number of vertices to find a diverse result set. The algorithms in the second set run the ranking function k times to select each result, and refine the search with some changes at each step. Although the former class of algorithms are preferred for practical use, they may not be able to reach to the intended diversity levels due to the highly greedy nature of the vertex selection process.

3.1. Diversification by vertex selection

The following approaches are used after getting the direction-aware relevancy (prestige) rankings of the vertices for a given set of seed nodes. The ranking function is selected as DARWR with parameters (κ, d).

Vertex-reinforced random walks (DIVRANK) [Mei et al. 2010]: For the random walk based methods mentioned in Section 2.3, the probabilities in the transition matrix $\mathbf{P}$ do not change over the iterations. Using vertex-reinforced random walk, DIVRANK adjusts the transition matrix based on the number of visits to the vertices. The original DIVRANK assumes that there is always an organic link for all the nodes returning back to the node itself with probability $(1-\alpha)$:

$$p_0(u,v) = \begin{cases} \alpha \frac{w(u,v)}{\delta(u)}, & \text{if } u \neq v \\ 1-\alpha, & \text{otherwise,} \end{cases} \tag{7}$$

where $w(u,v)$ is equal to 1 for $(u,v) \in E'$, and 0 otherwise. The transition matrix $\mathbf{P}_t$ at iteration t is computed with

$$\mathbf{P}_t(u,v) = (1-d)\, p^*(v) + d\, \frac{p_0(u,v)\, \eta_t(v)}{\sum_{z \in V} p_0(u,z)\, \eta_t(z)}, \tag{8}$$

where $\eta_t(v)$ is the number of visits of vertex v. It ensures that the highly ranked nodes collect more value over the iterations, resulting in the so called *rich-gets-richer* mechanism.

For each iteration of the defined vertex-reinforced random walk, the transition probabilities from a vertex u to its neighbors are adjusted by the number of times they are visited up to that point $\eta_t(v)$. Therefore, u gives a high portion of its rank to its frequently visited neighbors. Since the tracking of η_t is nontrivial, the authors propose to estimate it using two different models. One way is to employ the *cumulative ranks*, i.e., $\mathrm{E}[\eta_t(v)] \propto \sum_{i=0}^{t} \mathbf{p}_i(v)$, and since the ranks will converge after sufficient number of iterations, it can also be estimated with *pointwise ranks* as $\mathrm{E}[\eta_t(v)] \propto \mathbf{p}_t(v)$.

While adapting DIVRANK to our directional problem, we identified two problems: first, the initial ranks of all nodes should be set to a nonzero value; otherwise, the ranks cannot be distributed with (8) for both *pointwise* and *cumulative* estimation of η_t. Therefore, we set $p_0(v) = 1/n$ for all $v \in V$. Second, an organic link returning back to the node itself enables it to preserve its rank. This is problematic since p^* is only set for seed papers, and they tend to get richer over time. However, our objective is to distribute the probabilities over $V \setminus \mathcal{Q}$ to get a meaningful ranking. We solved this problem by removing the organic links of the seed papers, hence, distributing all of their ranks towards their neighbors instead of only α of them.

With the listed modifications, we propose the *direction-aware* DIVRANK algorithm using the transition probabilities

$$p_0'(u,v) = \begin{cases} 0, & \text{if } u \in \mathcal{Q}, u = v \\ \frac{(1-\kappa)}{\delta^+(u)}, & \text{if } u \in \mathcal{Q}, u \neq v, (u,v) \in E \\ \frac{\kappa}{\delta^-(u)}, & \text{if } u \in \mathcal{Q}, u \neq v, (v,u) \in E \\ (1-\alpha), & \text{if } u \notin \mathcal{Q}, u = v \\ \alpha \frac{(1-\kappa)}{\delta^+(u)}, & \text{if } u \notin \mathcal{Q}, u \neq v, (u,v) \in E \\ \alpha \frac{\kappa}{\delta^-(u)}, & \text{if } u \notin \mathcal{Q}, u \neq v, (v,u) \in E \end{cases} \tag{9}$$

which can be directly used in (8). Depending on the estimation method to be whether *cumulative* or *pointwise*, we refer to the *direction-aware* variants of the algorithm as

CDIVRANK and PDIVRANK, respectively.

Maximize the goodness measure (DRAGON) [Tong et al. 2011]: One of many diversity/relevance optimization functions found in the literature is the *goodness* measure. It is defined as:

$$f_{G'}(S) = 2\sum_{i\in S}\pi(i) - d\sum_{i,j\in S}\mathbf{A}'(j,i)\pi(j) - (1-d)\sum_{j\in S}\pi(j)\sum_{i\in S}p^*(i), \tag{10}$$

where $\mathbf{A}'$ is the row-normalized adjacency matrix of the graph. The original algorithm runs on the undirected citation graph G' and uses a greedy heuristic to find a near-optimal solution set. Accordingly, the *direction-aware goodness* measure f_G can be defined as:

$$f_G(S) = 2\sum_{i\in S}\pi(i) - d\kappa\sum_{i,j\in S}\mathbf{A}(j,i)\pi(j) - d(1-\kappa)\sum_{i,j\in S}\mathbf{A}(i,j)\pi(i), \tag{11}$$

where $\mathbf{A}$ is the row-normalized adjacency matrix based on directed graph, and the last part of (10) is always zero $\left(\sum_{i\in S}p^*(i)=0\right)$ since seed papers are never included in S. The direction-aware variant of the algorithm, running on the directed citation graph and using the ranking vector DARWR, is referred to as DRAGON.

Choose local maxima (LM): Because of the smoothing process of random walks, frequently visited nodes tend to increase the ranks of its neighbors [Mei et al. 2010]. Therefore, we propose to identify the papers that are local maxima and to return the top-k of them. This will guarantee that the nodes returned in this way are recommended by taking the smoothing process of random walks into account.

Once the ranks are computed, the straightforward approach to find the local maxima is to iterate over each node and check if its rank is greater than all of its neighbors' with an $\mathcal{O}(|E|)$ algorithm. In practice, one can mark all the vertices that have been identified as not being a local maxima to avoid traversing their adjacency list. This algorithm runs much faster since every rank comparison between two unmarked nodes (either local maxima or not) will mark one of them. The asymptotic complexity remains $\mathcal{O}(|E|)$. The LM algorithm is given in Algorithm 1.

Choose relaxed local maxima (γ-RLM): The drawback of diversifying with local maxima is that for large k's (i.e., $k > 10$), the results of the algorithm are generally no longer related to the queried seed papers, but some popular ones in unrelated fields, e.g., a set of well-cited physics papers can be returned for a computer science related query. Although this might improve the diversity, it hurts the relevance, hence, the results are no longer useful to the user.

In order to keep the results within a reasonable relevancy to the query and to diversify them, we relax the algorithm by incrementally getting local maxima within the top-γk results until $|S| = k$, and removing the selected vertices from the subgraph for the next local maxima selection. We refer to this algorithm as parameterized relaxed local maxima (γ-RLM) where γ is the relaxation parameter. Note that 1-RLM reduces to DARWR and ∞-RLM reduces to LM. The outline of the algorithm is given in Algorithm 2. In the experiments, we select $\gamma = k$ and refer this algorithm as k-RLM. In Section 4.5, we devise other experiments to see the effects of γ with respect to different measures.

ALGORITHM 1: Diversify with local maxima (LM)

```
Input: G' = (V, E'), π, k
Output: An ordered set of recommendations S
L ← empty list of (v, π_v)
for each v ∈ V do
    lm[v] ← LocalMax
for each v ∈ V do
    if lm[v] = LocalMax then
        for each v' ∈ adj[v] do
            if π_v' < π_v then
                lm[v'] ← NotLocalMax
            else
                lm[v] ← NotLocalMax
                break
        if lm[v] = LocalMax then
            L ← L ∪ {(v, π_v)}
PartialSort(L, k) w.r.t π_i non-increasing
S ← L[1..k].v, i.e., top-k vertices
return S
```

ALGORITHM 2: Diversify with relaxed local maxima (γ-RLM)

```
Input: G' = (V, E'), π, k, γ
Output: An ordered set of recommendations S
R ← PartialSort(V, γk) w.r.t. π_i non-increasing
R ← R[1 : γk]
while |S| < k do
    R' ← FindLocalMaxima(G, R, π)
    if |R'| > k − |S| then
        Sort(R') w.r.t. π_i non-increasing
        R' ← R'[1 : (k − |S|)]
    S ← S ∪ R'
    R ← R \ R'
return S
```

3.2. Diversification by query refinement

In this set of diversification algorithms, the ranking function is called multiple times while some of the parameters or graph structure are altered between those rankings.

Incremental ranking using absorbing random walks (GrassHopper) [Zhu et al. 2007]: GrassHopper is a well-known diversification algorithm which ranks the graph multiple times by turning at each iteration the highest-ranked vertex into a sink node[2]. Since the probabilities will be collected by the sink vertices when the random walk converges, the method estimates the ranks with the number of visits to each node before convergence.

The original method uses a matrix inversion to find the expected number of visits; however, inverting a sparse matrix makes it dense, which is not practical for the large and sparse citation graph we are using. Therefore, we estimate the number of visits

[2]A sink node only has a single outgoing edge to itself, so that all its rank stays trapped within the sink.

by iteratively computing the cumulative ranks of the nodes with DARWR.

Incremental ranking by graph sparsification (GSPARSE): In this algorithm, in contrast with GRASSHOPPER, after executing the ranking function, we propose to sparsify the graph by removing all the reference and citation edges around the highest ranked node and repeat the process until all k nodes are selected. Note that GRASSHOPPER converts the selected node into a sink node while GSPARSE disconnects it from the graph (see Alg. 3). This way, the vertices around the selected node becomes less connected, hence, they will attract less visits in a random walk.

ALGORITHM 3: Diversify by graph sparsification (GSPARSE)

Input: $G = (V, E)$, $\mathcal{Q}$, k
Output: An ordered set of recommendations S
$S \leftarrow \emptyset$
$G' \leftarrow G$
for $iter = 1 \rightarrow k$ **do**
  $ranks \leftarrow$ DARWR$(G' = (V', E'), \mathcal{Q})$
  $v \leftarrow \text{argmax}(ranks)$
  $S \leftarrow S \cup \{v\}$
  for each $v' \in adj[v]$ **do**
    $E' \leftarrow E' \setminus \{(v, v')\}$
  $V' \leftarrow V' \setminus \{v\}$
return S

4. EXPERIMENTS

4.1. Evaluation measures

We previously investigated the shortcomings of evaluating result diversification as a bicriteria optimization problem with a relevance measure that ignores diversity, and a diversity measure that ignores relevance to the query in [Kucuktunc et al. 2013]. Since the problem is similarly bicriteria, we argue that the relevance and diversity of the results should be evaluated with separate measures instead of a combined one.

Normalized relevance: The relevancy score of a set can be computed by comparing the original ranking scores of the resulting set with the top-k ranking list [Tong et al. 2011], defined as

$$\textit{rel}(S) = \frac{\sum_{v \in S} \pi_v}{\sum_{i=1}^{k} \hat{\pi}_i}, \tag{12}$$

where $\hat{\pi}$ is the sorted ranks in non-increasing order.

Difference ratio: The results of a diversity method are expected to be somewhat different than the top-k relevant set of results since, as our experiments will show, the set of nodes recommended by the original DARWR are not diverse enough. This is expected since highly ranked nodes will also increase the ranks of their neighbors [Mei et al. 2010]. Nevertheless, the original result set has the utmost relevancy. This fact can mislead the evaluation of the experimental results. Therefore, we decided to measure the difference of each result set from the set of original top-k nodes. Given the top-k relevant set $\hat{S}$, the difference ratio is computed with

$$\textit{diff}(S, \hat{S}) = 1 - \frac{|S \cap \hat{S}|}{|S|}. \tag{13}$$

Usefulness: The original ranking scores π actually show the usefulness of the nodes. Since these scores usually follow a power law distribution, the high ranked nodes collect most of the scores and the contribution of two low-ranked nodes to the *rel* measure can be almost the same even though the gap between their positions in the ranking is huge. Yet, the one with the slightly higher score might be useful where the other might not due to this gap. We propose the *usefulness* metric to capture what percentage of the results are actually *useful* regarding their position in the ranking:

$$use(S) = \frac{|\{v \in S : \pi_v \leq \tilde{\pi}\}|}{|S|}, \tag{14}$$

where $\tilde{\pi} = \hat{\pi}_{10\times k}$, i.e., the relevancy score of the node with rank $10 \times k$, for $k = |S|$, and $use(S)$ gives the ratio of the recommendations that are within top $10 \times k$ of the relevancy list.

ℓ-step graph density: A variant of graph density measure is the ℓ-step graph density [Tong et al. 2011], which takes the effect of indirect neighbors into account. It is computed with

$$dens_\ell(S) = \frac{\sum_{u,v\in S, u\neq v} d_\ell(u,v)}{|S| \times (|S|-1)}, \tag{15}$$

where $d_\ell(u,v) = 1$ when v is reachable from u within ℓ steps, i.e., $d(u,v) \leq \ell$, and 0 otherwise. The inverse of $dens_\ell(S)$ is used for the evaluation of diversity in [Mei et al. 2010].

ℓ-expansion ratio: Other diversity measures, the *expansion ratio* and its variant *ℓ-expansion ratio* [Li and Yu 2011] measure the coverage of the graph by the solution set. It is computed with

$$\sigma_\ell(S) = \frac{|N_\ell(S)|}{n}, \tag{16}$$

where $N_\ell(S) = S \cup \{v \in (V - S) : \exists u \in S, d(u,v) \leq \ell\}$ is the ℓ-step expansion set.

Goodness: direction aware alternative, given in (11).

Average year: The average publication year of the recommendation set.

Average pairwise distance: Pairwise shortest distance between the results is a measure of how connected or distant the recommendations are to each other. It is computed with

$$APD(S) = \frac{\sum_{u,v\in S, u\neq v} d(u,v)}{|S| \times (|S|-1)}. \tag{17}$$

Average MIN distance to $\mathcal{Q}$: Distance of the recommendations to the closest seed paper is a measure of relevance regarding the query:

$$AMD(S) = \frac{\sum_{v\in S} \min_{p\in\mathcal{Q}} d(s,p)}{|S|}. \tag{18}$$

Note that the intent-aware measures, such as α-normalized discounted cumulative gain (α-nDCG@k) [Clarke et al. 2008], intent-aware mean average precision (MAP-IA) [Agrawal et al. 2009], are not included to the discussions, but they are important measures for evaluating the diversity of the results when the data and queries have some already known categorical labels. Our problem has no assumptions of a known distribution that specifies the probability of an item belonging to a category.

As we list a number of measures, it is important to show that our experiments do not favor any group of measures that correlate with each other. Here, we investigate the listed measures (except the average publication year and runtime) by computing their pairwise correlations based on the results of the mentioned algorithms in Section 3.

Table II. Correlations of various measures. Pearson correlation scores are given on the lower triangle of the matrix. High correlations are highlighted.

	rel	*diff*	*use*	goodness	$dens_1$	$dens_2$	σ_1	σ_2	APD	AMD
rel	–									
diff	**-0.89**	–								
use	0.60	-0.59	–							
goodness	0.06	0.02	-0.21	–						
$dens_1$	-0.13	0.17	-0.48	0.16	–					
$dens_2$	-0.22	0.24	-0.32	0.05	0.58	–				
σ_1	-0.51	0.46	-0.75	0.07	0.14	0.32	–			
σ_2	-0.52	0.47	-0.73	0.04	0.13	0.32	**0.99**	–		
APD	0.27	-0.29	0.27	0.15	-0.47	-0.78	-0.27	-0.30	–	
AMD	-0.37	0.42	-0.78	0.51	0.52	0.24	0.47	0.43	-0.15	–

Table II shows the correlations of 10 measures as scatter plots as well as their correlation scores. For the graph diversity measures, ℓ-step expansion ratios (σ_1 and σ_2) are highly correlated among each other, showing that the reachable sets expand independent of the seed nodes (queries), and also proportional to a ratio, which is the average degree of the graph. On the other hand, none of the relevance or diversity measures has a high correlation with other measures.

4.2. Dataset collection and queries

We retrieved the metadata information on 2.2M computer science articles (as of May 2013) from DBLP[3], 830K technical reports on physics, mathematics, and computer science from arXiv[4], and 3M medical publications from PMC open access subset[5]. This data is well-formatted and disambiguated; however, it contains very few citation information (less than 470K edges). To increase the number of edges and inter-connect different disciplines, we imported the publications and reference relations from Cite-

[3]http://dblp.uni-trier.de/
[4]http://arxiv.org/
[5]http://www.ncbi.nlm.nih.gov/pmc/tools/openftlist/

Seer[6], ArnetMiner[7], and Related-Work project[8]. However, most of the data are automatically generated and often erroneous. We mapped each document to at most one document in each dataset with the title information (using an inverted index on title words and Levenshtein distance) and publication years. Using the disjoint sets, we merged the papers and their corresponding metadata from four datasets. The papers without any references or incoming citations are discarded. The final citation graph has about 11.4M papers and 33.1M directed edges, and will be used in the next version of our service.

The query set is composed of the actual queries submitted to the**advisor** service. We selected about 1840 queries where each query is a set $\mathcal{Q}$ of paper ids obtained from the bibliography files submitted by the users of the service who agreed to donating their queries for research purposes. $|\mathcal{Q}|$ varies between 1 and 697, with an average of 33.62.

4.3. Results

We run the algorithms on the described citation graph with varying k values (i.e., $k \in \{5, 10, 20, 50, 100\}$) and with the following parameters: α in (7) is selected as 0.25 as suggested in [Mei et al. 2010]. For the DARWR ranking, we use the default settings of the service, which are $d = 0.9$ for damping factor, and $\kappa = 0.75$ to get more recommendations from recent publications. In each run, the selected algorithm gives a set of recommendations S, where $S \subseteq V$, $|S| = k$, and $S \cap \mathcal{Q} = \emptyset$. The relevance and diversity measures are computed on S, and the average of each measure is displayed for different k values. The standard deviations are negligible, hence they are omitted.

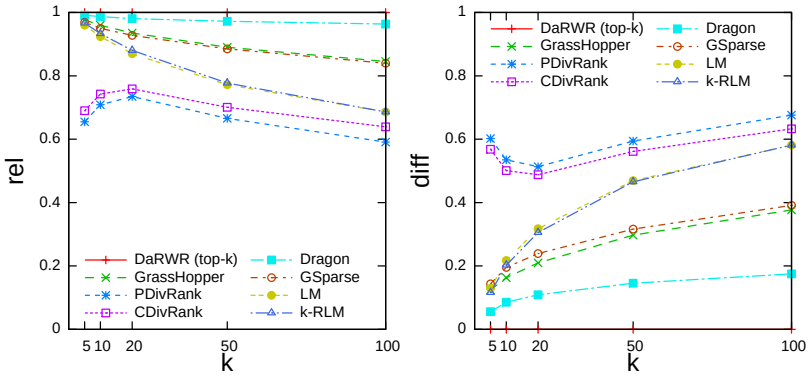


Fig. 2. Normalized relevance (left) and difference ratio (right) of the result set with respect to top-k results. Note that $rel_{\text{DARWR}} = 1$ and $diff_{\text{DARWR}} = 0$ since we compare the result set against itself. DRAGON returns almost the same result set as top-k.

Figure 2 shows the normalized relevancy and difference ratio of the recommendations compared to top-k results. It is arguable that a diversity-intended algorithm should maximize the relevancy since top-k results will always get the highest score, yet those have almost no value w.r.t. diversity. However, having a very low relevancy score indicates that the vertices have no connection to the query at all.

Since the normalized relevancy does not give us a clear idea of what is expected from those diversity-intended methods, we compare the set difference of the results

[6] http://citeseerx.ist.psu.edu/
[7] http://arnetminer.org/DBLP_Citation
[8] http://blog.related-work.net/data/

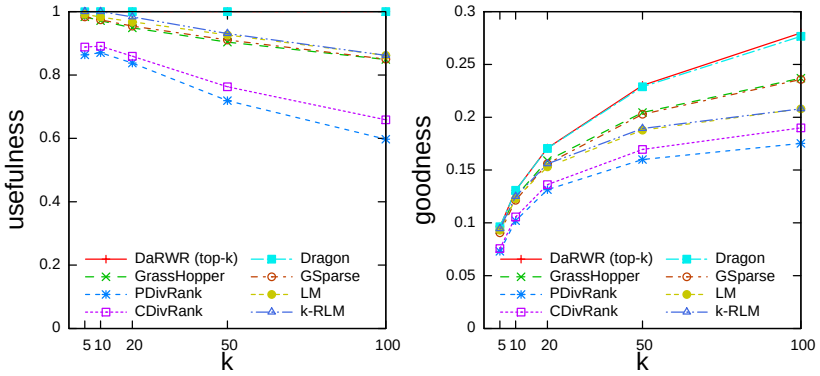


Fig. 3. Scores based on usefulness (left) and goodness (right) measures. DRAGON only slightly improves the goodness measure of the top-k results.

from top-k relevant recommendations. Figure 2-right shows that DRAGON gives a result set that is only 10-15% different than the top-k. In other words, the results of DRAGON differ in only one element when $k = 10$. DRAGON and the original top-k results score well on direction-aware goodness (Fig. 3-b); however, this also means that the goodness measure gives more importance to relevancy and less to diversity.

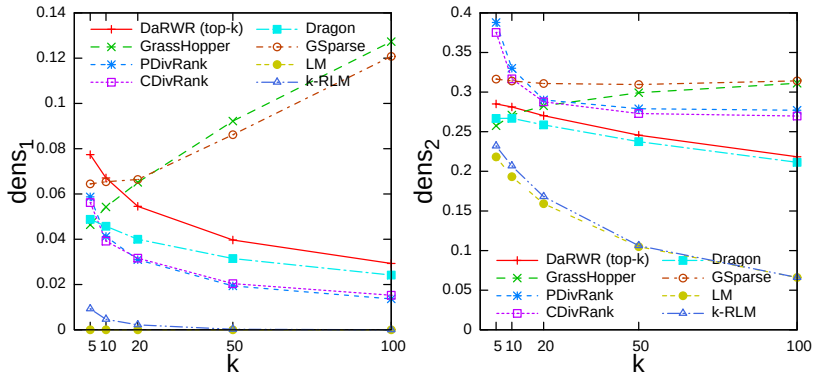


Fig. 4. ℓ-step graph density ($dens_\ell$) of the results. Note that $dens_1 \simeq 0$ for LM by construction. Both GRASSHOPPER and GSPARSE improve the diversity based on graph density for $k \leq 20$.

Graph density is frequently used as a diversity measure in the literature [Tong et al. 2011; Li and Yu 2011]. LM, k-RLM, and DIVRANK variants seem very promising (see Fig 4) for such a diversity objective. The same algorithms also perform good on ℓ-step expansion ratio (see Fig. 5), which is related to the coverage of the graph with the recommendations. GRASSHOPPER and GSPARSE perform worse in these diversity metrics. In particular, they are more dense than the results of DARWR.

After evaluating the results on various relevancy and diversity metrics, we are left with only a couple of methods that performed well on almost all of the measures: LM, k-RLM, and DIVRANK variants. However, Figure 6 shows that PDIVRANK and CDIVRANK give a set of results that are more connected (i.e., have a low average pairwise distance) and do not recommend recent publications (see Fig. 6-right) although κ is set

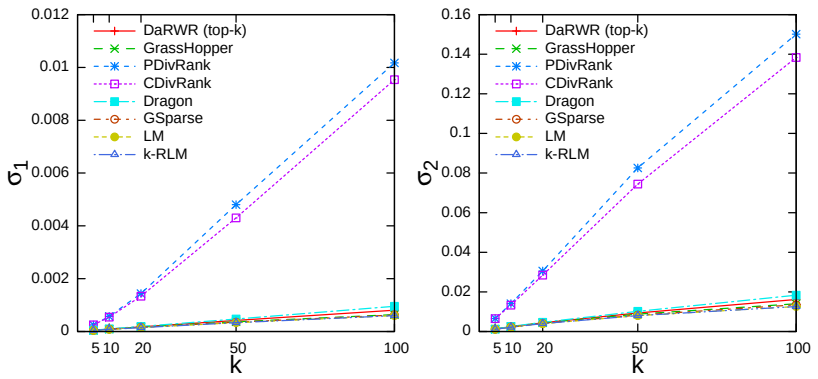


Fig. 5. ℓ-step expansion ratio (σ_ℓ) of the results. DIVRANK variants improve the diversity based on σ_ℓ.

accordingly. Since we are searching for an effective diversification method that runs on top of DARWR, DIVRANK variants are no longer good candidates.

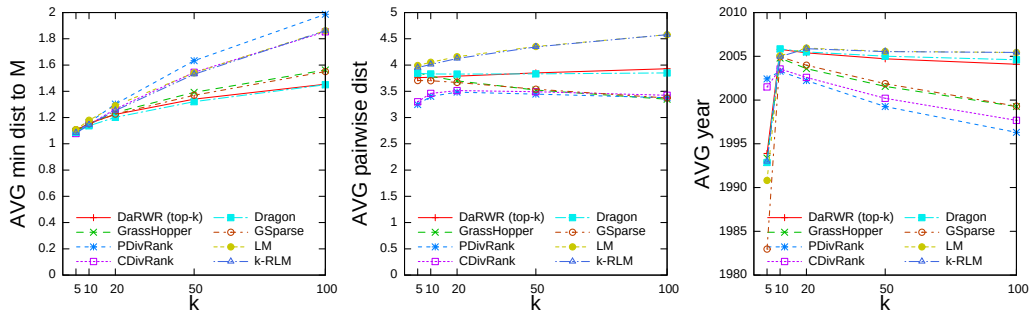


Fig. 6. Results based on average minimum distance to the query, average pairwise shortest distance between the recommended papers, and average publication year.

4.4. Scalability

The running time of the algorithms is also crucial for the web service since all the recommendations are computed in real-time. The experiments were run on the same architecture that the service is currently using. It has a 2.4GHz *AMD Opteron* CPU and 32GB of main memory. The CPU has 64KB L1 and 1MB L2 caches. The DARWR method and the dataset are optimized based on the techniques given in [Kucuktunc et al. 2012a]. In order to get a consistent runtime, the experiments are repeated ten times and averaged over these executions. Although the target architecture has 8 cores, the entire node was allocated for the experiment, but only one core was used.

It was expected that the complexity of the methods based on query refinement depend on and increase linearly with k. Figure 7 shows that GRASSHOPPER, and GSPARSE have the longest runtimes, even though they were faster than DIVRANK variants for $k \leq 10$. This behavior was previously mentioned in [Mei et al. 2010]. The running time of DRAGON is slightly higher than LM and k-RLM since it updates the goodness vector after finding each result.

In short, the query refinement-based methods (GRASSHOPPER, GSPARSE) have linearly increasing runtimes. DIVRANK variants require more iterations, therefore, more

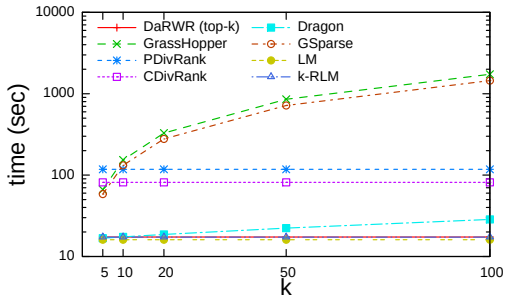


Fig. 7. Running times of the algorithms for varying k. DARWR, LM, k-RLM are equal.

time to converge. Finally, DRAGON, and especially LM and k-RLM are extremely efficient compared to other methods.

4.5. Parameter test

Our experiments on different relevance and diversity measures show that:

— DRAGON returns almost the same result set as top-k, while the graph density and expansion ratio measures also imply low diversity for their results,
— GRASSHOPPER and GSPARSE perform worse based on the diversity measures, and
— DIVRANK variants sacrifice direction-awareness for the sake of diversity,

whereas LM and k-RLM perform relatively good in almost all experiments, with a negligible computation cost on top of DARWR. k-RLM is slightly better than LM since it also improves the relevancy of the set to the query.

In order to understand the effects of the γ parameter to the quality of the result set, we display the results of γ-RLM with varying γ and k parameters in Figure 8. The experiments suggest that γ-RLM is able to sweep through the search space between all relevant (results of DARWR) and all diverse (results of LM) with a varying γ parameter. Therefore, this parameter can be set depending on the data and/or diversity requirements of the application.

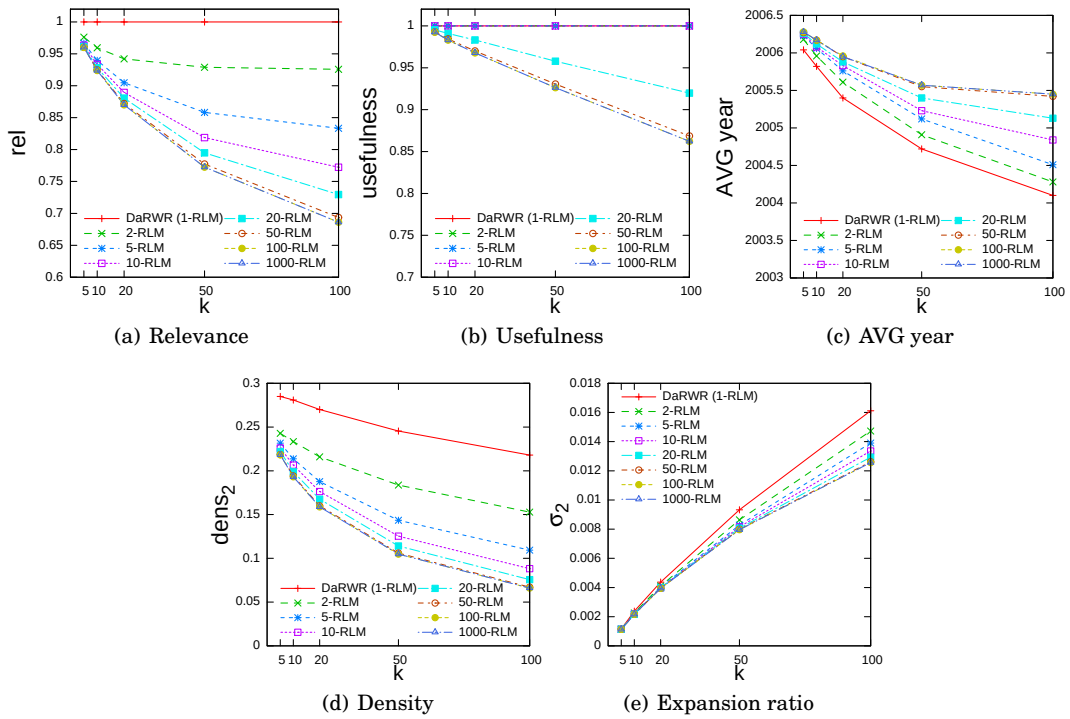
Figure 9 shows the results of γ-RLM with varying γ and κ parameters for $k = 20$. γ-RLM significantly improves the diversity of top-k results for any κ parameter. For $\gamma \geq 5$, average publication year of the results adapts better with the given κ, returning more recent papers as κ is closer to 1, and more traditional papers otherwise.

4.6. Intent-aware experiments

Here we present an evaluation of the intent-oblivious diversification algorithms against intent-aware measures. This evaluation provides a validation of the techniques with an external measure, such as *group coverage* [Li and Yu 2011] and *S-recall* [Zhai et al. 2003].

From the citation graph we obtain from different sources, we extract a subgraph of 545K vertices and 3.1M edges which corresponds to the citation graph of arXiv articles. We use this subgraph in intent-aware experiments because the authors of those articles assign at least one (e.g., “High Energy Physics - Phenomenology”, “Mathematics - Combinatorics”, “Computer Science - Computational Geometry”, etc.) out of 142 subjects. On average 1.52 subjects were assigned to each paper in the dataset.

The queries are selected with respect to the scenarios explained in [Kucuktunc et al. 2013]. Since our aim is to evaluate the results based on the coverage of different groups, we randomly generate 1000 query sets that represent multiple interests.



(a) Relevance (b) Usefulness (c) AVG year

(d) Density (e) Expansion ratio

Fig. 8. Parameter test on γ-RLM with varying γ and k parameters for $\kappa = 0.75$. As the method outputs more results with increasing k, the result set's relevance deteriorates and its diversity improves with increasing γ.

Specifically, for each query set, up to 10 random papers are selected from the citation graph as different interests of the user, and a total of 10 to 100 vertices within distance-2 of those interests are added to the query set. The intent of each query set $\mathcal{Q}$ is extracted by collecting the subjects of each seed node.

One measure we are interested in is the *group coverage* as a diversity measure [Li and Yu 2011]. It computes the number of groups covered by the result set and defined on subjects based on the intended level of granularity. However, this measure omits the actual intent of a query, assuming that the intent is given with the subjects of the seed nodes.

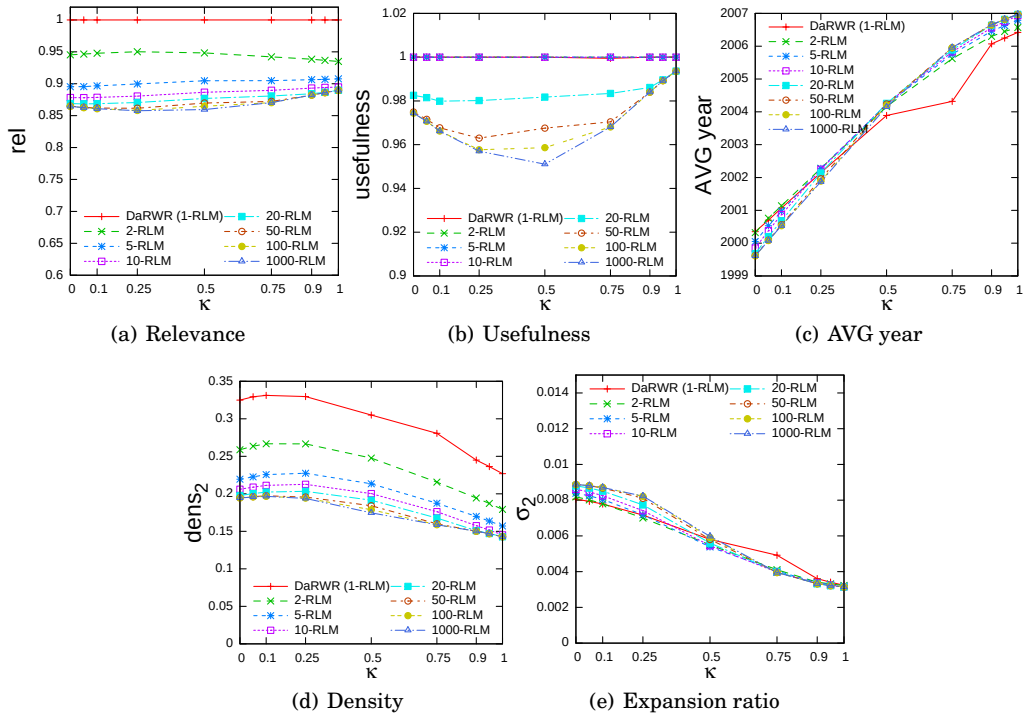
Subtopic recall (*S-recall*) has been defined as the percentage of relevant subjects covered by the result set [Zhai et al. 2003]. It has also been redefined as *Intent-Coverage* [Zhu et al. 2011], and used in the experiments of [Welch et al. 2011]. *S-recall* of a result set S based on the set of intents of the query I is computed with

$$S\text{-}recall(S, I) = \frac{1}{|I|} \sum_{i \in I} B_i(S), \tag{19}$$

where $B_i(S)$ is a binary variable indicating whether intent i is found in the results.

We give the results of group coverage and *S-recall* on subjects in Figure 10. The results of ALLRANDOM are included to give a comparison between the results of top-k relevant set (DARWR) and ones chosen randomly.

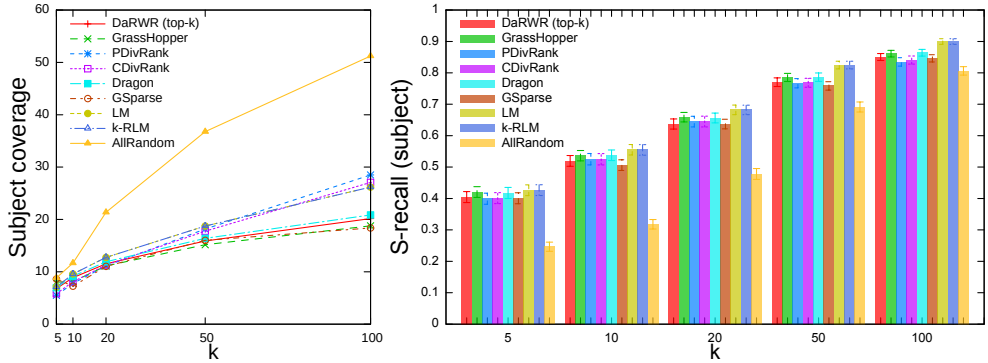
As the group coverage plots show, top-k ranked items of DARWR do not have the necessary diversity in the result set, hence, the number of groups that are covered by these



(a) Relevance (b) Usefulness (c) AVG year

(d) Density (e) Expansion ratio

Fig. 9. Parameter test on γ-RLM with varying γ and κ parameters for $k = 20$. γ-RLM significantly improves the diversity of the results. Average publication year of the results adapt better with the given κ for $\gamma \geq 5$.



(a) (b)

Fig. 10. Average intent-coverage and S-recall scores for the results of different diversification algorithms based on subjects. 95% confidence intervals for S-recall are also provided. For $k \geq 10$, LM and k-RLM are the only algorithms that have a significantly higher *S-recall* than DARWR, i.e., the confidence intervals do not intersect.

items are the lowest of all. On the other hand, a randomized method brings irrelevant items from the search space without considering their relevance to the user query. The results of all of the diversification algorithms reside between those two extremes,

where DIVRANK and LM variants cover the most, and GSPARSE and GRASSHOPPER cover the least number of groups.

However, *S-recall* index measures whether a covered group was actually useful or not. Obviously, ALLRANDOM scores the lowest as it dismisses the actual query (you may omit the *S-recall* on topics since there are only 6 groups in this granularity level). Among the algorithms, LM and k-RLM score the best overall while GRASSHOPPER have similar *S-recall* scores for $k = 10$ and 20, even though LM and k-RLM are much faster algorithms than GRASSHOPPER (cf. Figure 7).

4.7. Empirical results

Here, we try to exemplify the effects of diversifying recommendations with k-RLM method on a real world query[9]. The recommended and top-100 ranked papers are manually clustered and labeled into categories, i.e., graph mining (GM), generic SpMV (Sp), compression (C), multicore (MC), partitioning (P), GPU (GPU), and eigensolvers (E).

	top-k results		**k-RLM diversified**	
#	paper	label	paper	label
1	Govan09	GM	Govan09	GM
2	Kourtis08	C	Kourtis08	C
3	Lao10	GM	Lao10	GM
4	**Abbey10**	GM	Bradley10	GM
5	Bradley10	GM	Hoemmen10	Sp
6	Hoemmen10	Sp	**Saak64**	GPU
7	**Knight06**	GM	**Guo10**	GPU
8	**Davis97**	P	**Lee10**	MC
9	**Toledo97**	Sp	**Im04**	GPU
10	**Im00**	Sp	**Kaiser10**	MC

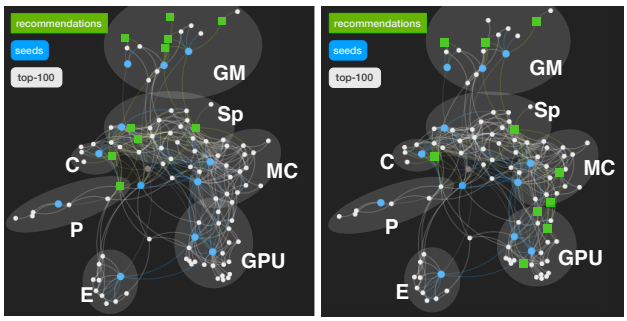


Fig. 11. top-10 and k-RLM diversified results for the given query (a), original and diversified recommendations are visualized with their categories (b,c). Diversified results bring about the same number of papers from categories that seed papers belong to.

The query is the bibliography of a submitted paper related to SpMV optimization for emerging architectures, hence a multidisciplinary paper. The query includes a couple of graph mining papers, and five out of ten relevance-only recommendations are related to graph mining, where three of them are neighbors. Figure 11 shows that the recommendations with k-RLM diversification improve the set of recommendations by eliminating redundant results and by covering other fields of interest. Indeed, no results from the multicore and GPU categories were returned before. After diversification, these two topics are covered. Moreover, the distribution of categories of k-RLM results resembles the one of the query, while top-k results do not.

5. CONCLUSIONS AND FUTURE WORK

In this work, we addressed the diversification of paper recommendations of the**advisor** service, which ranks the papers in the literature with a direction-aware personalized PageRank algorithm. While giving a survey of diversity methods designed specifically for random-walk-based rankings, we adapted those methods to our direction-aware problem, and proposed some new ones based on vertex selection and query refinement.

To evaluate the quality of the algorithms, we performed three types of experiments and established that the algorithm γ-RLM we proposed is best. First, using purely

[9]Available at http://theadvisor.osu.edu/csfeedback.php?q=e302d9fea1f22310cbf64c39a0a20d4e.ris,0.75

graph theoretic definitions of relevancy and diversity, we established that the algorithm γ-RLM exhibits good properties. It is fast. It returns relevant results significantly different from a relevant-only algorithm. And it minimizes the density of the graph induced by the result set. Second, we extracted a subset of the papers in our dataset which has been tagged with categorical informations by their authors. We established that γ-RLM is the algorithm that generates results that cover best the categories of the query. Finally, for a given query, we manually labeled the top-100 papers with topic information and verified that γ-RLM significantly improves the diversity of the returned papers compared to a relevance-only algorithm.

the**advisor** now uses γ-RLM to diversify the result set since it has performed best on both graph theoretic and categorical tests. However, the evaluation of diversification algorithms can be subjective. We plan to perform a user study to learn more about users' expectations. Also, textual information can allow to evaluate the diversity of the recommendations or be used to improve it.

REFERENCES

Rakesh Agrawal, Sreenivas Gollapudi, Alan Halverson, and Samuel Ieong. 2009. Diversifying search results. In *Proc. ACM Int'l Conf. Web Search and Data Mining*.

Sergey Brin and Lawrence Page. 1998. The anatomy of a large-scale hypertextual Web search engine. In *Proc. Int'l Conf. World Wide Web*.

Jaime Carbonell and Jade Goldstein. 1998. The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *Proc. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval*.

Ben Carterette. 2009. An Analysis of NP-Completeness in Novelty and Diversity Ranking. In *Proc. Int'l Conf. Theory of Information Retrieval*.

Charles L.A. Clarke, Maheedhar Kolla, Gordon V. Cormack, Olga Vechtomova, Azin Ashkan, Stefan Büttcher, and Ian MacKinnon. 2008. Novelty and diversity in information retrieval evaluation. In *Proc. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval*.

Marina Drosou and Evaggelia Pitoura. 2010. Search result diversification. *SIGMOD Rec.* 39, 1 (Sept. 2010).

Sreenivas Gollapudi and Aneesh Sharma. 2009. An axiomatic approach for result diversification. In *Proc. Int'l Conf. World Wide Web*.

Marco Gori and Augusto Pucci. 2006. Research Paper Recommender Systems: A Random-Walk Based Approach. In *Proc. WIC*.

Taher H. Haveliwala. 2002. Topic-sensitive PageRank. In *Proc. Int'l Conf. World Wide Web*.

M. M. Kessler. 1963. Bibliographic coupling between scientific papers. *American Documentation* 14 (1963), 10–25.

Onur Kucuktunc, Kamer Kaya, Erik Saule, and Umit V. Catalyurek. 2012a. Fast Recommendation on Bibliographic Networks. In *Proc. IEEE/ACM Int'l Conf. Advances in Social Networks Analysis and Mining*.

Onur Kucuktunc, Erik Saule, Kamer Kaya, and Umit V. Catalyurek. 2012b. Direction Awareness in Citation Recommendation. In *Proc. Int'l Workshop on Ranking in Databases (DBRank'12) in conjunction with VLDB'12*.

Onur Kucuktunc, Erik Saule, Kamer Kaya, and Umit V. Catalyurek. 2013. Diversified Recommendation on Graphs: Pitfalls, Measures, and Algorithms. In *Proc. Int'l Conf. World Wide Web*. 715–726.

Ni Lao and William Cohen. 2010. Relational retrieval using a combination of path-constrained random walks. *Machine Learning* 81 (2010), 53–67. Issue 1.

Steve Lawrence, C. Lee Giles, and Kurt Bollacker. 1999. Digital Libraries and Autonomous Citation Indexing. *Computer* 32 (1999), 67–71. Issue 6.

J Li and P Willett. 2009. ArticleRank: a PageRank-based alternative to numbers of citations for analyzing citation networks. *Proc. ASLIB* 61, 6 (2009).

Rong-Hua Li and J.X. Yu. 2011. Scalable Diversified Ranking on Large Graphs. In *Proc. IEEE Int'l Conf. Data Mining*.

Yicong Liang, Qing Li, and Tieyun Qian. 2011. Finding Relevant Papers Based on Citation Relations. In *Proc. Int'l Conf. Web-Age Information Management*.

David Liben-Nowell and Jon M. Kleinberg. 2007. The link-prediction problem for social networks. *JASIST* 58, 7 (2007), 1019–1031.

Bin Liu and H. V. Jagadish. 2009. Using trees to depict a forest. *Proc. VLDB Endowment* 2, 1 (2009).

Nan Ma, Jiancheng Guan, and Yi Zhao. 2008. Bringing PageRank to the citation analysis. *Inf. Process. Manage.* 44 (2008), 800–810. Issue 2.

Qiaozhu Mei, Jian Guo, and Dragomir Radev. 2010. DivRank: the interplay of prestige and diversity in information networks. In *Proc. ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining*.

H. P. F. Peters, Robert R. Braam, and Anthony F. J. van Raan. 1995. Cognitive Resemblance and Citation Relations in Chemical Engineering Publications. *Journal of the American Society for Information Science* 46, 1 (1995), 9–21.

Gerard Salton. 1963. Associative Document Retrieval Techniques Using Bibliographic Information. *J. ACM* 10 (1963), 440–457. Issue 4.

Henry Small. 1973. Co-citation in the scientific literature: A new measure of the relationship between two documents. *J. Am. Soc. Inf. Sci.* 24, 4 (1973), 265–269.

Trevor Strohman, W. Bruce Croft, and David Jensen. 2007. Recommending citations for academic papers. In *Proc. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval*.

Hanghang Tong, Christos Faloutsos, and Jia-Yu Pan. 2006. Fast Random Walk with Restart and Its Applications. In *Proc. IEEE Int'l Conf. Data Mining*.

Hanghang Tong, Jingrui He, Zhen Wen, Ravi Konuru, and Ching-Yung Lin. 2011. Diversified ranking on large graphs: an optimization viewpoint. In *Proc. ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining*.

Erik Vee, Utkarsh Srivastava, Jayavel Shanmugasundaram, Prashant Bhat, and Sihem Amer-Yahia. 2008. Efficient Computation of Diverse Query Results. In *Proc. IEEE Int'l Conf. Data Engineering*.

Michael J. Welch, Junghoo Cho, and Christopher Olston. 2011. Search result diversity for informational queries. In *Proc. Int'l Conf. World Wide Web*. 237–246.

Cheng Xiang Zhai, William W. Cohen, and John Lafferty. 2003. Beyond independent relevance: methods and evaluation metrics for subtopic retrieval. In *Proc. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval*. 10–17.

Xiaojin Zhu, Andrew B. Goldberg, Jurgen Van Gael, and David Andrzejewski. 2007. Improving Diversity in Ranking using Absorbing Random Walks. In *Proc. of HLT-NAACL*.

Xiaofei Zhu, Jiafeng Guo, Xueqi Cheng, Pan Du, and Hua-Wei Shen. 2011. A unified framework for recommending diverse and relevant queries. In *Proc. Int'l Conf. World Wide Web*. 37–46.

Cai-Nicolas Ziegler, Sean M. McNee, Joseph A. Konstan, and Georg Lausen. 2005. Improving recommendation lists through topic diversification. In *Proc. Int'l Conf. World Wide Web*.

Guido Zuccon, Leif Azzopardi, Dell Zhang, and Jun Wang. 2012. Top-k Retrieval Using Facility Location Analysis. In *Proc. European Conf. Information Retrieval*.

Received September 2012; revised July 2013; accepted –