Tag-based Paper Retrieval: Minimizing User Effort with Diversity Awareness

Nguyen Quoc Viet Hung, Do Son Thanh, Nguyen Thanh Tam, and Karl Aberer

École Polytechnique Fédérale de Lausanne
{quocviethung.nguyen,sonthanh.do,tam.nguyenthanh,karl.aberer}@epfl.ch

Abstract. As the number of scientific papers getting published is likely to soar, most of modern paper management systems (e.g. ScienceWise, Mendeley, CiteULike) support tag-based retrieval. In that, each paper is associated with a set of *tags*, allowing user to search for relevant papers by formulating tag-based queries against the system. One of the most critical issues in tag-based retrieval is that user often has difficulties in precisely formulating his information need. Addressing this issue, our paper tackles the problem of automatically suggesting new tags for user when he formulates a query. The set of tags are selected in such a way that resolves query ambiguity in two aspects: *informativeness* and *diversity*. While the former reduces user effort in finding the desired papers, the latter enhances the variety of information shown to user. Through studying theoretical properties of this problem, we propose a heuristic-based algorithm with several salient performance guarantees. We also demonstrate the efficiency of our approach through extensive experimentation using real-world datasets.

1 Introduction

With the rapid advances in science and technology, large collections of papers have been published every year. To manage such paper collections efficiently, many tag-based systems such as ScienceWise [1], Mendeley [3], and CiteULike [2] have been developed and received spectacular attentions. In these systems, each paper is associated with multiple tags, which often represent the domains it belongs to, the concepts it is related to, or the terms it contains. All associated tags in the repository are essential to enable tag-based retrieval that allows users to represent their search intents by choosing from a suggested list of tags and returns the relevant papers. For example, a user wants to retrieve the paper that he read before, but does not remember its name. He only has partial information about the paper (e.g. its domain and terms). By using the suggested tags, the user can easily figure out what he is exactly searching. As an another example, consider a user searching for papers of relevance to the research proposal he is working on. While the user is eventually interested in one or few papers, at the beginning he may have a lot of search queries in mind; thus a search with useful suggestion of tags is necessary to narrow down the choices. Motivated by these examples, we argue that tags can better help users specify their search intents rather than letting them issue the queries by themselves, especially if they do not know important keywords in the field.

In this work, we study the problem of minimizing user's effort in finding his expected paper(s) through an effective tag suggestion. More precisely, our goal is to minimize the expected number of tags which user need to put into the query. To the best of our knowledge, the closest work to ours is the research on query reformulation. In general, users are often not be able to state their search intents clearly when formulating a search query. The purpose of query reformulation is to provide additional information via query terms for users to reformulate their search intents. The terms are often ranked by different criteria such as co-occurrence patterns [20], latent topic model [5], and via knowledge bases [26]. The main difference between our work and the previous ones is that we rank the tags by their potential information towards reducing user effort.

The problem is challenging for several reasons. First, the dependencies between tags dynamically change according to the search context (i.e. current user query). Hence, it is necessary to develop a suggestion model that takes into account both the currently retrieved papers and the tags which were previously chosen into user query. Second, since the user's intent is not known until he is satisfied with the search, the problem of minimizing user effort cannot be solved in advance. As such, the suggestion needs to look-ahead possible choices by user when he formulates the next query, so that the user can reach the desired paper(s) with minimal (expected) number of querying steps. Third, there is a trade-off between information and diversity of the tags. Although suggesting the tags with high amount of information might improve the chances of reducing the search results quickly, user is also prevented from having a broad view of different domains on top of the suggestion.

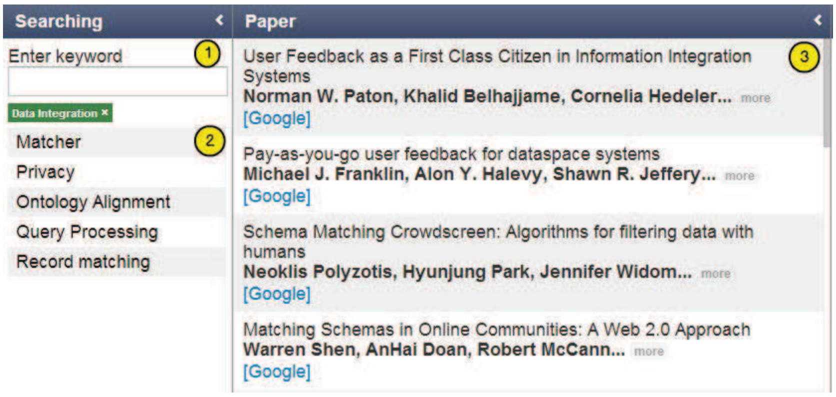
Addressing these challenges via a unified model of tag-based paper retrieval, this paper makes the following contributions.

- Section 2: We first provide a generic user interaction scheme for tag-based retrieval. Further, we introduce a formal model of the retrieval process. Then we motivate the requirements of tag suggestion.
- Section 3: We propose a goodness function that quantifies the quality of a tag suggestion solution by combining the two dimensions *informativeness* and *diversity* mentioned above. We also show that our function satisfies a set of useful properties.
- Section 4: We formulate the problem of finding a tag set with maximal goodness value. We prove that this problem is NP-hard. And thus, we propose a greedy algorithm with several salient performance guarantees to approximate the solution.

The remaining sections are structured as follows. Section 5 presents the experimental evaluations. Section 6 summarizes related work, before Section 7 concludes the paper.

2 Tag-based Paper Retrieval

User Interaction Scheme. Our tag-based paper retrieval framework implements a user interaction scheme as illustrated in Figure 1. Given a list of available tags in (2), a user chooses one of them to put into the query box (1). For this tag-based query, the system returns results as a set of papers in (3). Using tags as a query for retrieving papers helps user to narrow down the scope of research topics and quickly obtain the papers of interest. Moreover, he is also given an overview of all research topics, without spending any effort to rediscover these topics by manually reading the papers. In general, the result quality of tag-based search depends on how well the papers in the repository are annotated by tags. Our work is based on existing paper repositories, such as ScienceWise [1], Mendeley [3], and CiteULike [2], in which each paper is well-annotated with many meaningful tags by the experts in the field.



User Feedback as a First Class Citizen in Information Integration Systems
Norman W. Paton, Khalid Belhajjame, Cornelia Hedeler... more
[Google]

Pay-as-you-go user feedback for dataspace systems
Michael J. Franklin, Alon Y. Halevy, Shawn R. Jeffery... more
[Google]

Schema Matching Crowdscreen: Algorithms for filtering data with humans
Neoklis Polyzotis, Hyunjung Park, Jennifer Widom... more
[Google]

Matching Schemas in Online Communities: A Web 2.0 Approach
Warren Shen, AnHai Doan, Robert McCann... more
[Google]

Fig. 1: Tag-based Exploration User Interface

Tag-based Retrieval. We denote a repository of papers by $\mathcal{D}$, in which each paper $d \in \mathcal{D}$ is annotated by a set of tags T_d. We also denote $\mathcal{T} = \bigcup_{d \in \mathcal{D}} T_d$ as the set of all tags available in the system. Tag-based paper retrieval is the process of finding a paper (or papers) of interest through dynamically suggesting the tags of this paper (or these papers) to user. We assume that user does not know in advance exactly which paper he is looking for and which tags he should choose. Instead, he explores the repository by sequentially selecting the tags suggested by the system until he is satisfied with the search result. More precisely, we model the retrieval process as an interactive process, where in each step three actions are performed: (I) the system suggests a list of tags to user, (II) user chooses one of these suggested tags into the query, and (III) the system updates the set of retrieved papers of relevance to the chosen tags. In general, the retrieval set of papers is reduced after each step and the process ends once the retrieval goal is reached (e.g., user is satisfied). The main focus of our work is to suggest a good set of tags in each step such that the number of retrieved papers is reduced as fast as possible.

Technically, each user interaction step is characterized by a specific index i. Then $Q_i = \langle Q_i^+, Q_i^- \rangle$ denotes the tag-based query formulated by users in step i, where Q_i^+ contains a set of inclusive tags and Q_i^- contains a set of exclusive tags; i.e. $Q_i^+ \cap Q_i^- = \emptyset$, $Q_i^+, Q_i^- \subseteq \mathcal{T}$. For convenience, we denote the size of user query as $|Q_i| = |Q_i^+| + |Q_i^-|$. In the beginning, we have $Q_0 = \langle \emptyset, \emptyset \rangle$. Based on the query Q_{i-1} and the repository $\mathcal{P}$, the system suggests a list of tags $T_i \subseteq \mathcal{T} \setminus (Q_{i-1}^+ \cup Q_{i-1}^-)$, $|T_i| = k$ (action I). Among the suggested tags T_i, user chooses a particular tag t as either inclusion or exclusion into Q_i (action II). That is, $Q_i = \langle Q_{i-1}^+ \cup \{t\}, Q_{i-1}^- \rangle$ or $Q_i = \langle Q_{i-1}^-, Q_{i-1}^- \cup \{t\} \rangle$. In action III, the set of retrieved papers relevant to Q_i is denoted as D_i. A paper is considered relevant to a tag-based query if it contains all inclusive tags and does not contain any exclusive tags; i.e., $D_i = \{d \in \mathcal{D} \mid Q_i^+ \subseteq T_d \wedge Q_i^- \cap T_d = \emptyset\}$. A possible retrieval goal is that there remains only one paper or a set of papers sharing the same tags; i.e. $|D_i| = 1$ or $\forall d, d' \in D_i$, $T_d = T'_d$. Note that for brevity sake, we overload set notation for the suggestion list of tags T_i (or T), meaning that set operators applied to the list are evaluated based on the set of list elements.

Minimal User Effort with Diversity Awareness. In this work, we study the question of how to design a tag suggestion method that minimizes user effort with diversity awareness. In other words, the tags are ranked for suggestion along two dimensions:

- *Informativeness*: The tags are not independent; some tags always appear together in common papers while some others never go along with each other. Therefore, each tag has a distinguished amount of potential information. Suggesting the tags with higher potential information would provide more chances of minimizing the number of user interaction steps for retrieving the papers that truly match user intent.
- *Diversity*: The tags with high potential information might belong to the same domains, since they often have similar dependencies with the others. As such, only focusing on the *informativeness* dimension might prevent user from having a broad view of different domains. Therefore, there is a need of diversifying the list of tags suggested to user. In the absence of explicit knowledge about user intent, increasing the diversity (i.e. the number of domains) of the suggested tags would increase the probability of retrieving some papers that truly match the user's expectation.

To provide a unified quality measurement of tag suggestion, we propose a single comprehensive goodness function that combines both the informativeness degree and the diversity of the tags. The details are given in the next section.

3 Tag Suggestion Quality

In this section, we propose a quality measurement for tag suggestion. Given a user query Q_i and the set of retrieved papers D_i at step i, the quality of a tag set T_i is measured by a goodness function $g : 2^{\mathcal{T}} \rightarrow \mathbb{R}$, where $2^{\mathcal{T}}$ denotes the domain of possible tag sets. For brevity sake, we hereby omit the step index i of the notations (Q_i, T_i, etc.). The goodness value is composed of two notions: informativeness and diversity penalty. While the former reflects the degree of saving user effort of a tag when it is chosen by user, the latter addresses the diversity aspect by penalizing tags that are similar to each other.

3.1 Informativeness

As described above, user expresses his search intent by formulating a query from available tags. Based on the formulated query, our system retrieves a set of relevant papers. However, since user cannot often provide a concrete query that truly describes his search intent, the retrieved papers might not satisfy user expectation. In other words, there are always some degrees of uncertainty about matching user search intent with the retrieved papers. At the beginning of the retrieval process, this uncertainty is high since the query only has few tags and thus the set of retrieved papers is still broadened. During the course of the process, user incrementally refines his search intent by adding more tags into the query. When more tags are added, the set of retrieved papers is narrowed downed. Its uncertainty is continuously reduced until the query is specific enough to reflect user search intent.

Therefore, to minimize user effort (i.e. the number of tags needed to put into the query), we have to suggest the tags with the highest uncertainty reduction. For example, we have two currently retrieved papers p_1 and p_2, which are associated with the tag sets $\{t_1, t_2\}$ and $\{t_1, t_3\}$ respectively. User has three tags t_1, t_2, t_3 as options to formulate the next query. Consider two cases:

(i) User chooses t_1: the set of retrieved papers does not change since both p_1 and p_2 contain t_1. In other words, the uncertainty of the retrieved papers does not change. Suggesting t_1 has no benefit of reducing the uncertainty.

(ii) User chooses t_2 (or t_3): the number of retrieved papers reduces to only one (p_1 or p_2, for both inclusive and exclusive options). In other words, only one set of associated tags remains; i.e. the retrieved papers become certain or there is no uncertainty. Suggesting t_2 reduces the uncertainty.

Based on this observation, we introduce the concept of *informativeness*, which measures the amount of uncertainty reduction of a tag when it is chosen into user query (e.g. informativeness of t_1 is 0, of t_2 is > 0). Suggesting the tags with low informativeness (low information gain) like t_1 requires user to choose many tags, while suggesting the tags with high informativeness (high information gain) like t_2 or t_3 makes the retrieval process faster. Hence, to minimize user effort, we should suggest the tags with high informativeness. In the following, we propose a probabilistic formulation to compute the informativeness of a tag.

Probability of a tag. As mentioned earlier, we denote D as the set of retrieved papers given a user query Q. The probability that a particular tag t is used in D then becomes:

$$p_t = \frac{|\{d \in D | t \in T_d\}|}{|D|} \tag{1}$$

Recall that T_d is the set of tags annotated with the paper d. The probability distribution of all tags available in the retrieval is thus denoted as $\Omega(D) = \{p_t | t \in \mathcal{T}\}$. Intuitively, tags that appear in all papers have probability of 1; whereas, tags that do not appear in the same papers with the tags in user query have probability of 0.

Uncertainty of matching user intent. We compute the uncertainty of matching user intent of a set of retrieved papers D as the Shannon entropy over the probability distribution of the tags:

$$H(D) = -\sum_{p_t \in \Omega(D)} [p_t \log p_t + (1 - p_t) \log(1 - p_t)] \tag{2}$$

where $H(D) \geq 0$. A set of papers in which each paper is annotated with different sets of tags implies a high uncertainty and vice-versa. The more user effort (i.e. more tags are added to the query), the lower value of the uncertainty. As a consequence, the retrieval process ends when the uncertainty reaches zero. Indeed, $H(D) = 0$ means that all the associated tags have probability equal to either 0 or 1. In other words, all the retrieved papers are annotated with an identical set of tags, which converges to user search intent.

Conditional uncertainty. We now compute the uncertainty of the retrieved papers if user chooses a particular tag. Since the choice of regarding t as inclusive tag or exclusive tag in the query is not known before-hand, the conditional uncertainty should be measured as the expected amount across both cases. Formally, we define the conditional uncertainty w.r.t a particular tag as the entropy conditioned on that tag:

$$H(D|t) = p_t \times H(D^{+t}) + (1 - p_t) \times H(D^{-t}) \tag{3}$$

where $p_t \in \Omega(D)$ is the probability that t is used in D as aforementioned. $D^{+t} = \{d \in D | t \in T_d\}$ and $D^{-t} = \{d \in D | t \notin T_d\}$ are respectively the set of retrieved papers after the inclusiveness and exclusiveness of t in user query.

Informativeness computation. We compute the informativeness of a tag t following a decision theoretic approach, cf. [28]. More precisely, we measure the amount of uncertainty reduction obtaining by the decision that t is selected; i.e. this reduction is computed as the difference between the ambiguity of the retrieved papers before and after user selects t. Formally, we have:

$$IG(t) = H(D) - H(D|t) \tag{4}$$

With a normalized form ($\in [0, 1]$) as:

$$h(t) = \frac{IG(t)}{\max_{t' \in T_D} IG(t')} \tag{5}$$

Any tag with informativeness equal to zero would have no contribution to reduce the uncertainty. The more informativeness of the tag, the more chances of the uncertainty being improved. In the sense of user effort, we should suggest the high informative tags to reduce the number of user interaction steps. Moreover, it is worth noting that at the beginning of the retrieval process, the query is empty. In this case, we consider $h(t) = 1$ for every tag, implying that all of them are initially considered equal for the suggestion.

3.2 Diversity Penalty

Computing informativeness helps us select more minimum-effort driven tags into the suggestion list. However, the most informative tags often belong to many common papers, resulting in a redundant suggestion. To increase the variety of the suggestion list, we penalize the tags that are similar to each other. The similarity between two tags reflects the amount of information that is shared in their common papers. While the computation of tag similarity is given at the end of this section, we first formulate the notion of diversity penalty. The idea is that the more similar between the tags, the higher amount of penalty is applied. Technically, the diversity penalty of a set of suggested tags T is calculated in terms of the pair-wise similarity between the tags weighted by their informativeness:

$$\phi(T) = \sum_{t,t' \in T} h(t) S(t, t') h(t') \tag{6}$$

where $S(t, t') \in [0, 1]$ is the similarity score between any two tags t and t' (the more similar, the higher value). We weight the tag similarity by informativeness of the tags to penalize similar tags with high informativeness more than those with low informativeness. This is motivated by the need to allow more chances of selecting dissimilar tags (despite of lower informativeness) to increase the diversity of the suggestion.

Similarity Computation. The similarity between tags should depend on user query and thus be computed dynamically during the paper retrieval process. For example, we consider two scenarios: (i) user query is “data mining”, (ii) user query contains “data mining” and “clustering”. In the first scenario, the two tags “DBSCAN” and “k-means” are similar since they are one of many well-known techniques in the field of data mining. In the second scenario, since "DBSCAN" and "k-mean" are the name of two different approaches in the clustering topic, they are dissimilar. Therefore, we propose a query-based probabilistic measurement for tag similarity (or dissimilarity) as follows.

Given a user query Q, the dissimilarity between two tags t_1, t_2 can be measured by the KL divergence [35] of two probability distributions when user chooses either t_1 or t_2:

$$\xi(t_1 \| t_2) = \sum_t p_t^1 \log \frac{p_t^1}{p_t^2} \tag{7}$$

where $p_t^1 \in \Omega(D_1)$ and $p_t^2 \in \Omega(D_2)$ are the probability of a tag t when either t_1 or t_2 is chosen. That is, $D_1 = \{d \in D | t_1 \in T_d\}$ and $D_2 = \{d \in D | t_2 \in T_d\}$. Since there is no meaningful notion of order in similarity, we use a commonly used symmetric variation:

$$\xi'(t_1, t_2) = \xi(t_1 \| t_2) + \xi(t_2 \| t_1) \tag{8}$$

However, the KL divergence still does not take into account the relationship between the two tags and user query Q. For example, for the tag set T = {"data mining"}, we could add t_1 = "shared memory" and t_2 = "message passing" whose meanings are not related to "data mining". To improve this, we weight the KL divergence by the conditional probabilities of the two tags and therefore discount additional tags that have no real relation with the query. As a result, the tag dissimilarity can be defined as:

$$\xi''(t_1, t_2) = p_{t_1} p_{t_2} \xi'(t_1, t_2) \tag{9}$$

where $p_{t_1}, p_{t_2} \in \Omega(D)$. With a further normalization (into $[0, 1]$) and inversion of dissimilarity, we have the final form of tag similarity:

$$S(t_1, t_2) = 1 - \frac{\xi''(t_1, t_2)}{\max_{i,j} \xi''(t_i, t_j)} \tag{10}$$

In general, the larger similarity between two given tags, the higher penalty they receive (i.e. the higher chance they are not selected). The aim of diversifying the tag suggestion becomes the selection of the tags that are sufficiently dissimilar with each other.

3.3 Put It Altogether

To balance informativeness and diversity in a top-k selection of tags, we design a quality measure for such a selection. On the one hand, the goodness measure should incorporate given informative scores of tags in a fine-grained level, by weighting the importance of tags unequally. The idea behind is that tags stemming from a large group of similar tags are often associated with popular papers, which implies a high chance to satisfy user information needs. On the other hand, the goodness measure should penalize similar tags. This is motivated by the need to increase diversity in the suggestion.

Our goodness measure for a selection of tags T is based on the overall, weighted informativeness of a selected tag, which is reduced by the diversity penalty of the tags that have also been selected. Intuitively, this approach favors tags from big clusters of similar tags, but penalizes the selection of multiple informative tags that are very similar to each other. Technically, given T_D as the set of tags associated with the current set of retrieved papers D, we define $q(t) = \sum_{t' \in T_D} S(t, t') \cdot h(t')$ as the importance of tag $t \in T_D$. With $w \in \mathbb{R}^+$ as a positive weight parameter, our goodness measure is defined as follows:

$$g(T) = w \sum_{t \in T} q(t) h(t) - \phi(T) \tag{11}$$

The proposed notion of goodness satisfies the following properties [10], whose proofs can be found in the appendix. First, our notion of goodness shows monotonicity; i.e., when adding more tags to an existing selection, the goodness of the overall selection will increase.

Proposition 1 (Monotonicity) *Let T_D be a set of tags associated with a particular paper set D. For any $w \geq 2$, $\forall T_1, T_2 \subseteq T_D, T_1 \cap T_2 = \emptyset$, we have $g(T_1 \cup T_2) \geq g(T_1)$.*

Second, our goodness measure shows submodularity, which refers to the property that marginal gains in goodness start to diminish due to saturation of the objective. That is, the marginal benefit of adding tags to the selection decreases w.r.t. the size of the selection.

Proposition 2 (Submodularity) *Let T_D be a set of tags associated with a particular paper set D. For any $w > 0$, $\forall T \subseteq T_D, t_1, t_2 \in T_D \setminus T$, we have $g(T \cup \{t_1\}) + g(T \cup \{t_2\}) \geq g(T \cup \{t_1, t_2\}) + g(T)$.*

4 Efficient Tag Suggestion

In this section, we first formulate our tag suggestion problem. Due to the NP-hardness of the problem, we then propose a greedy algorithm. After that, we prove various performance guarantees for the proposed algorithm.

4.1 Problem Definition

Using the notion of goodness, we define tag suggestion as an optimization problem. That is, we are interested in finding a selection of top-k tags that maximize the goodness measure:

Problem 1 (Tag Suggestion) *Let T_D be a set of tags associated with the retrieved papers and k be a threshold for the number of tags. Then, the tag suggestion problem is defined to be:*

$$\underset{T \subseteq T_D, |T| = k}{\operatorname{argmax}} \; g(T) \tag{12}$$

Here, selection of the top-k tags is of particular practical relevance for information retrieval, cf., [23]. An appropriate value for k depends on the user and the application context. In general, the problem of tag suggestion turns out to be NP-Complete, whose proof can be found in the appendix.

Theorem 1 *The k-tag suggestion problem is NP-Complete.*

Algorithm 1: Heuristic algorithm for tag suggestion.

```
   input  : A set of tags $T_D$ associated with the retrieved papers,
            a weight factor $w \geq 2$, and a threshold for the number of tags $k$.
   output : A selection of tags $T^* = \langle t_1, \dots, t_k \rangle, t_i \in T_D, 1 \leq i \leq k$.
1  $T^* \leftarrow \emptyset$ ;
   // Compute ranking score for each tag
2  Let $r : T_D \rightarrow \mathbb{R}, r(t) \mapsto w \cdot h(t) \cdot \sum_{t' \in T_D} S(t, t')h(t')$;
3  while $|T^*| < k$ do
4      $t_m \leftarrow \mathrm{argmax}_{t \in T_D, t \notin T^*} r(t)$ ;
5      $T^* \leftarrow T^* \cap \{t_m\}$ ;
       // Update ranking score for the remaining tags
6      $r' \leftarrow r$;
7      Let $r : T_D \rightarrow \mathbb{R}, r(t) \mapsto r'(t) - 2 \cdot h(t_m) \cdot S(t, t_m) \cdot h(t)$;
8  return $T^*$
```

4.2 Algorithm

Given the complexity of the tag suggestion problem, we now present a heuristic algorithm to approximate its optimal solution [10]. The main idea of our algorithm is to start from the null set and add one element at a time, taking at each step the element which increases the goodness of the suggestion list most. To achieve a provably near-optimal solution, our algorithm exploits the two aforementioned properties of the goodness function g, i.e., monotonicity ad submodularity. In essence, the algorithm iteratively expands the selection of tags by adding the tag that maximizes the goodness value, thus it can be bounded. Solving the problem requires k iterations.

The details of our heuristic are given in Algorithm 1. It takes a set of tags T_D, a weight factor w, and a threshold for the number of tags k as input and returns a selection T^* of k tags. We begin by computing a ranking score for each tag $t \in T_D$ that is based on the weight factor, the tag informativeness, and the tag importance (line 2). In the actual greedy selection step, we select k tags. In each iteration, we add the tag with the highest ranking score (lines 4 and 5), before the ranking score is updated for the remaining tags (line 7). The latter avoids re-computation of the ranking scores from scratch in each iteration. As mentioned above, we overload set notation for the suggestion list of tags T^* for brevity sake. When presented in user interface, the tags are listed top-down in the decreasing order of ranking score (i.e. from left to right of the sequence representation).

4.3 Algorithm Analysis

The proposed algorithm shows several desirable properties. First, the approximation error is bounded.

Guarantee 1 (Near-Optimality) *Algorithm 1 is a (1- 1/e)-approximation for the tag suggestion problem.*

Proof. For any monotone, submodular function f with $f(\emptyset) = 0$, it is known that an iterative algorithm selecting the element e with maximal value of $f(I \cup \{e\}) - f(I)$ with I as the set of elements selected so far has a performance guarantee of $(1 - 1/e) \approx 0.63$ [24]. This result is applicable to algorithm 1, since our goodness function g is monotonic (proposition 1) and submodular (proposition 2), it holds $g(\emptyset) = 0$ (eq. (11)), and the ranking score is defined as $r(t) = g(T^* \cup \{t\}) - g(T^*)$ (lines 2 and 7).

Next, we consider the complexity of our heuristic.

Guarantee 2 (Complexity) *The time complexity and the space complexity of Algorithm 1 are* $O(|T_D|^2 + k|T_D|)$ *and* $O(|T_D|)$*, respectively.*

Proof. Time complexity: The quadratic term $|T_D|^2$ stems from the computation of the ranking score. The linear term $k|T_D|$ is explained by k iterations, in each of which we iterate over all remaining tags, for selection of t_{max} and for updating the ranking score. Space complexity: Storing tag similarities requires $\frac{|T_D||T_D-1|}{2}$ space since S is symmetric and $S(t, t)$ is fixed.

Further, our algorithm shows stability in the selection, which is important to support multi-resolution (i.e. in cases user wants to see more tags in the suggestion list). For example, if a user is first presented with the top-10 tags, but then extends the suggestion list to the top-20, the expectation is clearly that the top-10 remain unchanged.

Guarantee 3 (Stability) *For* T^* *as returned by algorithm 1, let* $T^*_{k_1} = \langle t_1, \ldots, t_{k_1} \rangle, T^*_{k_2} = \langle t'_1, \ldots, t'_{k_2} \rangle$ *be selections with* $t_i \in T^*$*,* $1 \leq i \leq k_1$*,* $t'_j \in T^*$*,* $1 \leq j \leq k_2$*, and* $0 < k_1 \leq k_2$*. Then, it holds that* $t_i = t'_i$ *for* $1 \leq i \leq k_1$*.*

Proof. In Algorithm 1, the construction of T^* is performed stepwise and elements are never removed from T^*. Moreover, the selection is deterministic: we always add a new tag with the highest ranking score (line 4). Thus, the larger selection sequence comprises the smaller selection sequence as a prefix.

5 Experiments

This section presents a comprehensive experimental evaluation to verify the effectiveness of our tag-based paper retrieval framework. In particular, we first discuss the experimental setup including datasets and evaluation measures. Then, we proceed to report the following experiments: (i) evaluations on informativeness, and (ii) evaluations on diversity. The results highlight that the proposed tag suggestion algorithm performs well in terms of both user effort and diversity aspect.

5.1 Experimental Settings

Dataset. Our prototype is developed on top of the ScienceWise platform since it supports API to retrieve data and has a rich tag collection. The ScienceWise's data contains 16725 scientific papers and 15083 tags. Each paper has 70 tags in average. The ScienceWise platform itself has not supported tag suggestion in the search results yet.

Evaluation Measures. For comparative evaluation, we study the following measures.

Domain Coverage. This metric measures the diversity of a top-k list of tags in terms of coverage of domains. It indicates the proportion of possible domains (which might be of interest to user) the tag list can capture. Formally, we run k-meloids clustering to divide the set of all available tags T_D into k clusters, based on the tag similarity proposed in Section 3.2. The domain $dom(t)$ of a tag t is the cluster it belongs to. The domain coverage ($\in [0, 1]$) of top-k tag suggestion T^* is defined as the number of domains of its tags over the total number of domains:

$$DC(T^*) = \frac{|\bigcup_{t \in T^*} dom(t)|}{k} \tag{13}$$

Normalized Informativeness. This metric measures the informativeness of the tag suggestion list with respect to the top-k tags with highest informativeness; i.e., it indicates how well the informativeness of the tags is preserved when diversity is taken into account. Formally, the normalized informativeness ($\in [0, 1]$) of top-k tag suggestion T^* from the set of candidate tags T_D is defined as the sum of their informativeness scores over the sum of the k highest informativeness scores:

$$nH(T_D, T^*) = \frac{\sum_{t \in T^*} h(t)}{\max_{T \subseteq T_D, |T|=|T^*|} \sum_{t \in T} h(t)} \tag{14}$$

User Effort. To quantify the amount of time user spends to retrieve the desired papers, we compute the user effort as the number of interaction steps of the retrieval process described in Section 2. Each interactive step is counted when user selects a new tag to be added into the query. Formally, we have:

$$E = |Q^+| + |Q^-| \tag{15}$$

5.2 Evaluations on Informativeness

The goal of this evaluation is to verify the soundness of the proposed informativeness function of a tag. To this end, we will study the informativeness in two aspects: (i) paper amount reduction – how many retrieved papers are reduced after user chooses a tag, and (ii) user effort – how many tags user need to choose in the retrieval.

Informativeness vs. Paper Amount Reduction. In this experiment, we only consider one user interaction step of the retrieval process. We assume that user is interested in a particular paper, which is associated with a set of tags. The user query is simulated by randomly choosing some of these tags. Given a simulated query, we rank the tags by the decreasing order of informativeness. For each of the top-10 tags, we put it into the query as inclusive if it is contained in the tag set and exclusive otherwise. Then we retrieve the papers of the new query and measure the amount reduction of retrieved papers.

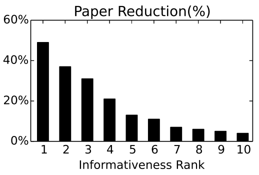


Fig. 2: Query Size = 0

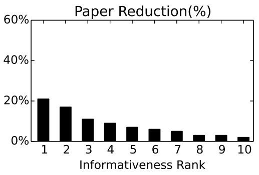


Fig. 3: Query Size = 10

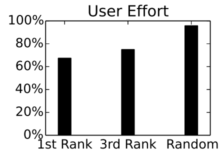


Fig. 4: Informativeness vs. User Effort

Figure 2 and Figure 3 illustrate the results for different query sizes (size = 0 and size = 10). The report numbers are averaged over 100 different targeted papers (these papers have more than 15 tags). The X-axis is the rank of tags in terms of informativeness. The Y-axis is the relative reduction of the amount of retrieved papers. An interesting finding is that the higher rank of tags, the more papers are reduced. For example, the tag with highest informativeness (rank 1) gives about 50% reduction, whereas the tag with lowest informativeness (rank 10) gives less than 5% reduction. This supports the soundness of our informativeness function in capturing user effort. Another noticeable observation is that as more tags are selected into the query, the number reduction of retrieved papers is smaller. For example, with query size = 0, the reduction of the rank-1 tag is about 50%, while this number is only about 20% with query size = 10. This is reasonable because after each user interaction step, the set of retrieved papers and the set of their associated tags are narrowed down. As such, the percentage of papers sharing the common tags is higher and selecting these tags would return mostly the same papers.

Informativeness vs. User Effort. In this experiment, we simulate the whole retrieval process. Like the previous experiment, we assume that user is interested in a particular paper, which is associated with a set of tags. At the beginning, we initialize user query by randomly choosing one of these tags. In each interaction step, user receives a suggested tag and put it into the query (the tag is regarded as inclusive or exclusive based on the target paper). The process stops when only one paper remains or all the remaining papers share the same set of tags. Three tag suggestion strategies are studied: (i) *1st Rank* – suggest the tag with highest informativeness, (ii) *3rd Rank* – suggest the tag of rank-3 in the decreasing order of informativeness, (iii) *Random* – suggests a random tag to user.

Figure 4 depicts the result, which is averaged over 100 simulations (i.e. 100 different target paper). The X-axis is the tag suggestion strategy. The Y-axis is the percentage of user effort over the number of tags contained in the target paper. A key finding is that the *Random* strategy incurs most user effort (95.73%). This can be explained by the fact that with *Random* strategy, user has to go through many redundant tags, which do not (or rarely) reduce the number of papers. Another interesting observation is that the more informativeness of the tag, the more user effort is reduced. Indeed, the *1st Rank* strategy takes the least user effort (67.38%), whereas the *3rd Rank* strategy requires more user effort (74.94%). This supports that user effort can be reflected through the informativeness of a tag. Suggesting the tag with the highest informativeness does indeed reduce user effort the most.

5.3 Evaluations on Diversity

In this experiment, we would like to verify the soundness of the diversity aspect of tag suggestion. More precisely, we will compare two tag suggestion strategies: (1) with diversity: the suggestion list of tags is computed by the proposed algorithm, (2) without diversity: the suggestion list of tags is computed by returning the top tags with highest informativeness values. For the strategy 1, we randomly set the tunning parameter w (trade-off between informativeness and diversity) according to uniform distributions $\mathscr{U}(0,1)$ and $\mathscr{U}(1,2)$, respectively. The final numbers are computed as the average over 100 runs. We vary the number of suggested tags k from 5 to 55, and compare the two strategies according to different aspects as follows.

Figure 5 illustrates the results on the diversity aspect. A key finding is that strategy 1 is always better than strategy 2 in terms of domain coverage. For example, while

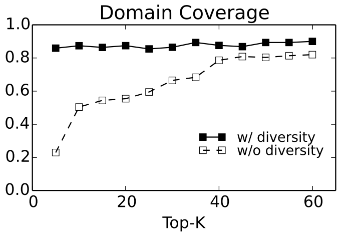


Fig. 5: Diversity

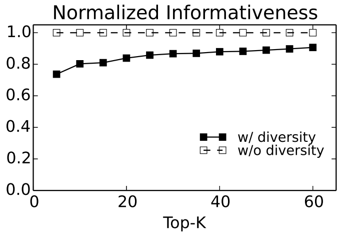


Fig. 6: Informativeness

the domain coverage of strategy 1 is always greater than 0.8, the domain coverage of strategy 2 is only about 0.2 with k = 5. This supports the fact that our proposed algorithm performs well in producing a diverse list of suggested tag. Another noticeable observation is that the difference of domain coverage between the two strategies is smaller when *k* increases. For instance, with *k* = 60, the domain coverage of strategy 2 is nearly 0.8. This is because when *k* is higher, strategy 2 will include more tags with lower informativeness, up to the point that all tags in the list are dissimilar enough among themselves, resulting in high domain coverage.

Figure 6 presents the result on the informativeness aspect. By definition, the normalized informativeness of strategy 2 is always equal to 1. An interesting finding is that the normalized informativeness of strategy 1 is not much lower than strategy 2 in comparison with the domain coverage. For example, with *k* = 5, the difference of domain coverage is more than 0.6 while the difference of normalized informativeness is less than 0.25. This implies that in spite of producing a diverse list of suggested tags, our proposed algorithm still keeps most of the informativeness amount of the tags. In other words, the tags with high informativeness values are preserved, which goes beyond the trade-off between diversity and informativeness. Another important observation is that the normalized informativeness of strategy 1 increases when the suggestion size is higher. This is reasonable since the tags that are diverse often have different values of informativeness. When the number of suggested tags increases, our algorithm will add both the tags with high informativeness and the tags with low informativeness, up to the point that the two strategies share most of common tags with each other.

6 Related Work

Our work aims to reduce user effort for retrieving relevant papers in tag-based paper management platforms. It is mainly related to tag-based retrieval, query suggestion, and diversification, which are briefly reviewed as follows.

Tag-based Retrieval. In the last decades, there has been an increasing development of tag-based retrieval systems, which allow to add tags (manually or automatically) to existing resources such as images and videos. The research efforts in tag-based retrieval can be broadly categorized into three types, namely *annotating*, *ranking*, and *presenting*. *Annotating* involves determining the set of tags best describing a resource [16,35,39,36,34]. *Ranking* aims to compute a relevance score between a query and a resource [22,19,13]. *Presenting* focuses on improving user satisfaction by effectively presenting the tags or search results to users [18,37,33].

Tag-based retrieval for scientific papers is a distinguished and recognized direction. This is because using textual search on research articles has some limitations, for example, full-text access is not always available [6] and OCR errors are inherently found [30]. Moreover, different from other resources (web pages, image, videos), scientific papers are associated with much more tags since there is a lot of scientific concepts across different domains (e.g. in our dataset, each paper is associated with 70 tags in average). This distinct characteristic opens up an opportunity to design more complex mechanisms by exploiting potential information of the tag collection. A wide range of tag-based paper retrieval systems have been developed with reliable and high-quality tags such as ScienceWise [1], Mendeley [3], and CiteULike [2]. Moreover, there is also a considerable number of research outcomes on this direction, including tag-based search engine [14], semantic-based framework [27], and collaborative tagging [25].

Query Suggestion. Query suggestion (a.k.a. query reformulation, query expansion, query completion) is a supportive method to improve search productivity. In general, users are often not be able to state their search intents clearly when formulating a search query. The purpose of query suggestion is to provide additional information for users to help them reformulate their queries. In the literature, query suggestion has been studied in different contexts. In [5], the authors exploited query log of the search engine to suggest new query terms for the current user query. Instead of using query log, the authors of [20] made use of existing keywords provided by social annotation services to generate and rank the new queries for suggestion. In the same line, the authors of [26] extracted candidate query terms from existing Wikipedia articles related to user query. In the context of image search, the work in [37] uses representative images for user to look ahead the search results of query terms.

Diversification. The diversification problem has been long acknowledged in information retrieval [9,18]. It aims to improve user satisfaction by providing a diverse view of information, thereby increasing the probability of returning some information that truly matches the user's expectation. Various applications that have benefited from diversification include sentiment analysis [4], web search [12], database search [7], large-scale visualization [31], social network [40] and recommender systems [11]. In our case, since users cannot often precisely and exhaustively describe their queries, increasing diversity of tag-query suggestion will provide users more chances to find the desired papers quickly. We propose a *function-based* approach [17] for tag diversification, which is "less heuristic" than the *threshold-based* [32] and the *graph-based* [38] approaches.

To summarize, our work differs from previous research in the following aspects: (1) we do not aim to provide an "auto-complete" feature like the previous works. Rather, we study a different aspect of query suggestion with the goal of minimizing user's effort in retrieving the information that truly matches his search intent. (2) we jointly consider user effort minimization and diversification by designing a comprehensive goodness function, which guides the on-the-fly computation of suggested tags according to the current user query. Moreover, it is worth noting that although the proposed algorithm is demonstrated on the context of paper retrieval, it can be applied for other domains such as document retrieval and image retrieval. It should be also emphasized that our work is not about tagging online contents [16] (i.e. *Annotating*. Instead, we leverage the generated tags to better support the retrieval of these contents (which cannot be accessed via textual search).

7 Conclusions and Future Work

This work proposes a novel approach that enables tag-based retrieval in online archives of scientific papers. To make these archives searchable, each paper is associated with a set of pre-defined descriptive keywords, so-called tags. We study the problem of how to efficient suggest new tags for user to formulate his query intent. The goal is to not only reduce the efforts of user in reaching his search intent, but also increase the diversity of the suggested tags. In particular, we define the notion of goodness measure that captures both the informativeness and diversity aspects of the tags. Based on this measure, we formulate the tag suggestion problem as the identification of a set of k tags with maximal goodness value. Through studying theoretical properties of this problem, we propose a heuristic-based algorithm with several salient performance guarantees. Finally, we present a comprehensive experimental evaluation indicating that the approach allows for effective and efficient retrieval of real-world scientific data.

Our work opens up several future research directions. First, the proposed quality measurement can be used to evaluate existing query suggestion methods, especially the user-effort aspect. Second, we can investigate other dimensions to be considered in the quality measurement. Third, this paper focuses on searching for scientific papers, yet, our tag-based retrieval framework (in particular the tag suggestion algorithm) can be applied for a variety of domains, such as business documents and social medias. Fourth, although the suggested tags are presented as a list in our context, we can also study other presentation options such as hierarchical and categorical-like structures. Fifth, our work could be tailored to take into account the meta-data (e.g. citation [21]), if available, of the scientific papers to further refine their relevance (not only based on tags). When each paper has multiple search dimensions, we can develop more sophisticated cost models [15] as well. Moreover, one can also improve the retrieval performance by relevance feedback [29], which is out of the scope of this paper.

Acknowledgment The research has received funding from the EU-FP7 EINS project (grant number 288021) and the ScienceWise project.

References

1. http://sciencewise.info
2. http://www.citeulike.org/
3. http://www.mendeley.com/
4. Aktolga, E., Allan, J.: Sentiment diversification with different biases. In: SIGIR. pp. 593–602 (2013)
5. Bing, L., Lam, W., Wong, T.L.: Using query log and social tagging to refine queries based on latent topics. In: CIKM. pp. 583–592 (2011)
6. Cohen, A.M., Hersh, W.R.: A survey of current work in biomedical text mining. Briefings in bioinformatics pp. 57–71 (2005)
7. Drosou, M., Pitoura, E.: Disc diversity: Result diversification based on dissimilarity and coverage. In: PVLDB. pp. 13–24 (2012)
8. Feige, U., Peleg, D., Kortsarz, G.: The dense k-subgraph problem. Algorithmica pp. 410–421 (2001)
9. Goffman, W.: A searching procedure for information retrieval. ISR pp. 73–78 (1964)

10. He, J., Tong, H., Mei, Q., Szymanski, B.: Gender: A generic diversified ranking algorithm. In: NIPS. pp. 1142–1150 (2012)
11. Hurley, N., Zhang, M.: Novelty and diversity in top-n recommendation – analysis and evaluation. TOIT pp. 1–30 (2011)
12. Iwata, M., Sakai, T., Yamamoto, T., Chen, Y., Liu, Y., Wen, J.R., Nishio, S.: Aspectiles: Tile-based visualization of diversified web search results. In: SIGIR. pp. 85–94 (2012)
13. Jain, V., Varma, M.: Learning to re-rank: Query-dependent image re-ranking using click data. In: WWW. pp. 277–286 (2011)
14. Jomsri, P., Sanguansintukul, S., Choochaiwattana, W.: A comparison of search engine using ”tag title and abstract” with citeulike - an initial evaluation. In: ICITST. pp. 1–5 (2009)
15. Kashyap, A., Hristidis, V., Petropoulos, M.: Facetor: cost-driven exploration of faceted query results. In: CIKM. pp. 719–728 (2010)
16. Kim, J.W., Candan, K.S., Tatemura, J.: Organization and tagging of blog and news entries based on content reuse. J Sign Process Syst pp. 407–421 (2010)
17. Küçüktunç, O., Saule, E., Kaya, K., Çatalyürek, U.V.: Diversified recommendation on graphs: pitfalls, measures, and algorithms. In: WWW. pp. 715–726 (2013)
18. van Leuken, R.H., Garcia, L., Olivares, X., van Zwol, R.: Visual diversification of image search results. In: WWW. pp. 341–350 (2009)
19. Li, X., Snoek, C.G.M., Worring, M.: Learning social tag relevance by neighbor voting. In: TMM. pp. 1310–1322 (2009)
20. Lin, Y., Lin, H., Jin, S., Ye, Z.: Social annotation in query expansion: A machine learning approach. In: SIGIR. pp. 405–414 (2011)
21. MacRoberts, M.H., MacRoberts, B.R.: Problems of citation analysis: A critical review. JASIST pp. 342–349 (1989)
22. Maniu, S., Cautis, B.: Network-aware search in social tagging applications: Instance optimality versus efficiency. In: CIKM. pp. 939–948 (2013)
23. Manning, C.D., Raghavan, P., Schütze, H.: Introduction to information retrieval, vol. 1. Cambridge university press (2008)
24. Nemhauser, G., Wolsey, L., Fisher, M.: An analysis of approximations for maximizing submodular set functions–i. MP pp. 265–294 (1978)
25. Noël, S., Beale, R.: Sharing vocabularies: Tag usage in citeulike. In: BCS-HCI. pp. 71–74 (2008)
26. Oliveira, V., Gomes, G., Belém, F., Brandão, W., Almeida, J., Ziviani, N., Gonçalves, M.: Automatic query expansion based on tag recommendation. In: CIKM. pp. 1985–1989 (2012)
27. Prokofyev, R., Boyarsky, A., Ruchayskiy, O., Aberer, K., Demartini, G., Cudré-Mauroux, P.: Tag recommendation for large-scale ontology-based information systems. In: ISWC. pp. 325–336 (2012)
28. Russell, S.J., Norvig, P., Canny, J.F., Malik, J.M., Edwards, D.D.: Artificial intelligence: a modern approach, vol. 74. Prentice hall Englewood Cliffs (1995)
29. Salton, G., Buckley, C.: Improving retrieval performance by relevance feedback. JASIST (1997)
30. Sebastiani, F.: Machine learning in automated text categorization. CSUR pp. 1–47 (2002)
31. Skoutas, D., Alrifai, M.: Tag clouds revisited. In: CIKM. pp. 221–230 (2011)
32. Vieira, M.R., Razente, H.L., Barioni, M.C.N., Hadjieleftheriou, M., Srivastava, D., Traina, C., Tsotras, V.J.: On query result diversification. In: ICDE. pp. 1163–1174 (2011)
33. Wang, M., Yang, K., Hua, X.S., Zhang, H.J.: Towards a relevant and diverse search of social images. In: TMM. pp. 829–842 (2010)
34. Wang, Q., Ruan, L., Zhang, Z., Si, L.: Learning compact hashing codes for efficient tag completion and prediction. In: CIKM. pp. 1789–1794 (2013)
35. Weinberger, K.Q., Slaney, M., Van Zwol, R.: Resolving tag ambiguity. In: MM. pp. 111–120 (2008)

36. Xie, L., He, X.: Picture tags and world knowledge: Learning tag relations from visual semantic sources. In: MM. pp. 967–976 (2013)
37. Zha, Z.J., Yang, L., Mei, T., Wang, M., Wang, Z.: Visual query suggestion. In: MM. pp. 15–24 (2009)
38. Zhang, B., Li, H., Liu, Y., Ji, L., Xi, W., Fan, W., Chen, Z., Ma, W.Y.: Improving web search results using affinity graph. In: SIGIR. pp. 504–511 (2005)
39. Zhu, G., Yan, S., Ma, Y.: Image tag refinement towards low-rank, content-tag prior and error sparsity. In: MM. pp. 461–470 (2010)
40. Zhu, X., Goldberg, A.B., Van Gael, J., Andrzejewski, D.: Improving diversity in ranking using absorbing random walks. In: HLT-NAACL. pp. 97–104 (2007)

Appendix - Proofs

NP-Complete. We prove Theorem 1 by reduction to the Densest k-Subgraph problem, which is known to be NP-Complete [10,8]. Let $G = (V, E)$ be an undirected graph with vertices V and edges E. Let W be the $|V| \times |V|$ binary connectivity matrix (symmetric), i.e., $W_{i,j} = 1$ if $\{i, j\} \in E$, and $W_{i,j} = 0$ otherwise. Then, the Densest k-Subgraph problem requires identifying a subgraph of k vertices with a maximal number of edges:

$$\operatorname*{argmax}_{\hat{V}\subseteq V, |\hat{V}|=k} \sum_{i,j\in\hat{V}} W_{i,j}$$

which is equivalent to

$$\operatorname*{argmax}_{I=(V\setminus\hat{V}), |\hat{V}|=k} 2 \sum_{i\in\hat{V}, j\in I} W'_{i,j} + \sum_{i,j\in I} W'_{i,j} \tag{16}$$

where $W'_{i,j} = 1 - W_{i,j}$. Now we will show that eq. (16) can be viewed as an instance of the optimization problem in eq. (12). To this end, let all informative scores be one ($h(t) = 1$ for all $t \in T_D$) and choose $w = 2$. Then, our objective function $g(T)$ becomes:

$$\begin{aligned} g(T) = 2\sum_{t\in T} q(t) - \sum_{t_1,t_2\in T} S(t_1,t_2) = 2\sum_{t_1\in T}\sum_{t_2\in T_D} S(t_1,t_2) - 2\sum_{t_1,t_2\in T} S(t_1,t_2) + \sum_{t_1,t_2\in T} S(t_1,t_2) \\ = 2\sum_{t_1\in(T_D\setminus T)}\sum_{t_2\in T} S(t_1,t_2) + \sum_{t_1,t_2\in T} S(t_1,t_2) \end{aligned} \tag{17}$$

The latter is equivalent to the objective function in eq. (16), so that selection of k tags corresponds to the finding the densest subgraph of $(|V| - k)$ nodes.

Monotonicity. With $w \geq 2$, we have:

$$g(T_1 \cup T_2) - g(T_1) = w\sum_{t\in T_2} q(t)h(t) - (\sum_{t\in T_2, t'\in T_1} h(t)S(t,t')h(t') + \sum_{t\in T_1, t'\in T_2} h(t)S(t,t')h(t') + \sum_{t,t'\in T_2} h(t)S(t,t')h(t'))$$

$$= w\sum_{t\in T_2} h(t)\sum_{t'\in T_D} S(t,t')h(t') - (2\sum_{t\in T_1, t'\in T_2} h(t)S(t,t')h(t') + \sum_{t,t'\in T_2} h(t)S(t,t')h(t')) \geq 2\sum_{t\in T_2} h(t)\sum_{t'\in T_D} S(t,t')h(t') -$$

$$(2\sum_{t\in T_1, t'\in T_2} h(t)S(t,t')h(t') + \sum_{t,t'\in T_2} h(t)S(t,t')h(t')) = 2\sum_{t\in T_2}(\sum_{t'\in T_D} S(t,t')h(t') - \sum_{t'\in T_1\cup T_2} S(t,t')h(t')) = 2\sum_{t\in T_2}\sum_{t'\notin T_1\cup T_2} S(t,t')h(t') \geq 0$$

which completes the proof of monotonicity.

Submodularity. From eq. (11), we have:

$$g(T \cup \{x\}) - g(T) = wq(x)h(x) - 2h(x)\sum_{t\in T} S(x,t)h(t) + h^2(x) \tag{18}$$

Following eq. (18), we have:

$$g(T \cup \{t_1\}) + g(T \cup \{t_2\}) \geq g(T \cup \{t_1, t_2\}) + g(T) \Leftrightarrow g(T \cup \{t_1\}) - g(T) \geq g(T \cup \{t_2\} \cup \{t_1\}) - g(T \cup \{t_2\})$$

$$\Leftrightarrow wq(t_1)h(t_1) - 2h(t_1)\sum_{t\in T} h(t)S(t,t_1) + h^2(t_1) \geq wq(t_1)h(t_1) - 2h(t_1)\sum_{t\in T\cup\{t_2\}} h(t)S(t,t_1) + h^2(t_1) \Leftrightarrow 2h(t_1)h(t_2)S(t_1,t_2) \geq 0$$

which completes the proof of submodularity.