FolkDiffusion: A Graph-Based Tag Suggestion Method for Folksonomies

Zhiyuan Liu[1,3,4], Chuan Shi[2,3,4], and Maosong Sun[1,3,4]

[1] Department of Computer Science and Technology, Tsinghua University
[2] Department of Electronic Engineering, Tsinghua University
[3] State Key Lab on Intelligent Technology and Systems, Tsinghua University
[4] National Lab for Information Science and Technology, Tsinghua University
lzy.thu@gmail.com, shichuanwuhan@gmail.com, sms@tsinghua.edu.cn

Abstract. Collaborative social tagging is a popular and convenient way to organize web resources. All tags compose into a semantic structure named as folksonomies. Automatic tag suggestions can ease tagging activities of users. Various methods have been proposed for tag suggestions, which are roughly categorized into two approaches: content-based and graph-based. In this paper we present a heat diffusion method, i.e., FolkDiffusion, to rank tags for tag suggestions. Compared to existing graph-based methods, FolkDiffusion can suggest user- and resource-specific tags and prevent from topic drift. Experiments on real online social tagging datasets show the efficiency and effectiveness of FolkDiffusion compared to existing graph-based methods.

Keywords: Folksonomies, tag suggestions, FolkDiffusion, heat diffusion.

1 Introduction

Social tagging is a popular way to organize and share web resources such as documents, bookmarks and photos. Resources, tags and users are three essential parts in a social tagging system, in which users assign tags to resources and share with others. Social tagging systems can be categorized according to the type of resources. For example, Flickr (www.flickr.com) is designed to share photos; CiteULike (www.citeulike.org) and BibSonomy (www.bibsonomy.org) is to share academic references; and del.icio.us (www.delicious.com) is to share bookmarks. A common feature of these tagging systems is they all provide tag suggestions to ease the process of social tagging by recommending tags for a user given a resource.

Many methods have been proposed for tag suggestions. For the resources with rich text contents such as blogs and product reviews, content-based methods are proved to be effective [14]. However, for the resources with less text contents but having rich collaborative information, graph-based methods turn out to be useful. In graph-based methods, users, resources and tags are grouped into a tripartite graph, where there are connections only between users, resources and tags, and there are no connections within users, resources and tags themselves. After that, some ranking algorithms, such as PageRank [12], are performed on the graph. By biasing the ranking values to the given user and resource, we can rank tags specific to the given user and resource. FolkRank is a

representative graph-based method based on PageRank on user-resource-tag tripartite graph. FolkRank, proposed as a tag suggestion method in [3], was shown to outperform other content-free methods including collaborative filtering.

FolkRank, however, suffers from the problem of topic drift. Starting from the given user and resource, although FolkRank makes PageRank biased to the given user and resource to a certain extent, it always finally jumps into globally popular tags when reaching stationary state. This is caused by the characteristic of PageRank, which was originally proposed to rank web pages according to their link topology in isolation.

To solve this issue, in this paper we present a new graph-based ranking method, i.e., FolkDiffusion, for tag suggestions. FolkDiffusion is motivated by heat diffusion process. Heat diffusion is a physical phenomenon. The basic theory of heat diffusion is that heat always flows from the position with high temperature to the position with low temperature. Recently, the idea of heat diffusion is introduced in various applications such as dimension reduction [1], classification [6], anti-spam of web pages [17], social network analysis [9] and search query suggestion [8]. In these work, the input data is considered to be a medium and the heat is diffused from a given set of start points to other data points. After iterations of heat diffusion, the heat values of data points indicate their relatedness with the initial start points.

In this paper, we consider tag suggestions as a task of finding most related tags with the given user u and resource r. Firstly, we construct a graph with users, resources and tags, in which edge weights represent the relatedness between them. Based on the idea of heat diffusion, we initially set u and r having the temperature higher than zero. Then we set all tags, users and resources except u and r to be zero temperature. After that, the heat begins to flow from u and r on the graph according to the edges between them. The edge between two vertices can be imagined as the pipe of an air-conditioner for heat flow, and the edge weight indicates the diffusion speed of heat. After several iterations of diffusion, the heat values on tags indicate their relatedness with the original user u and resource r. The most related tags are suggested for u and r.

The important characteristic of FolkDiffusion is that the heat only flows from high temperature to low temperature. We can imagine the process as follows. FolkDiffusion firstly recognizes a group of most related objects. The initially recognized related objects are then endowed with the power to decide which may be further related. This feature prevents FolkDiffusion from topic drift, and guarantees the suggested tags are topically related to the given user and resource. In experiments, we will demonstrate the performance of FolkDiffusion compared with other graph-based methods including FolkRank.

2   Related Work

For resources with rich text contents, many content-based methods have been proposed for tag suggestions. Ohkura et al. [11] proposed a Support Vector Machine-based tag suggestion system. Katakis et al. [5] used a hierarchical multi-label text classifier to find the proper tags for a document. Mishne [10] used a search-based nearest neighbor method to suggest tags, where the tags of a new document were collected from the most relevant documents in the training set. Lipczak et al. [7] extracted keywords from the title of a document, and filtered them with a user's used tags to get final suggestions. Tatu

et al. [15] combined tags from similar documents and extracted keywords to provide tag suggestions. We find that most of content-based methods regard tag suggestions as either a text classification problem or a keyword extraction problem. In experiments of this paper, we also demonstrate a simple but effective content-based method, i.e., kNN, as a baseline method for tag suggestions.

Another class of tag suggestion systems is based on the connections between users, tags and resources, which does not take the content of resources into consideration and is thus named as graph-based methods. Xu et al. [16] used collaborative filtering to suggest tags for bookmarks. Jaschke et al. [3] proposed FolkRank, a PageRank-like iterative algorithm to find the most related tags for the given resource and user. PageRank was originally used for ranking web pages according to the topology of web graph. In PageRank we can set preference values to a subset of pages to make the final ranking values biased to these selected pages. In fact, FolkRank computes the relatedness between tags and the specific user and resource by setting the given user and resource to high preference values in PageRank.

3 Graph-Based Tag Suggestions

3.1 Problem Formulation

In social tagging systems, a folksonomy is defined as a tuple $\mathbb{F} = (U, T, R, Y)$, where U is the set of users, T is the set of tags and R is the set of resources. Y is a ternary relation between U, T and R, $Y \subseteq U \times T \times R$. A tuple record $(u, r, t) \in Y$ is called a tag assignment, which means user u assigned tag t to resource r. The goal of tag suggestions is to predict a set of tags $\{t\}$ for a given pair of user and resource (u, r).

The basic idea of graph-based tag suggestions is to construct a graph with users, resources and tags as vertices and build edges according to user tagging behaviors. After building the graph, we can adopt some graph-based ranking algorithms to rank tags for a specific user and resource. Then the top-ranked tags are recommended to users.

To describe the graph-based methods more clearly, we give some mathematical notations. For the folksonomy $\mathbb{F} = (U, T, R, Y)$, we firstly convert it into an undirected tripartite graph $G_\mathbb{F} = (V, E)$. In $G_\mathbb{F}$, the vertices consist of users, resources and tags, i.e., $V = U \cup R \cup T$. For each tagging behavior of user u assigning tag t to resource r, we will add edges between u , r and t, i.e., $E = \{\{u, r\}, \{u, t\}, \{r, t\}|(u, t, r) \in Y\}$.

In $G_\mathbb{F}$, we have the set of vertices $V = \{v_1, v_2, \cdots, v_N\}$ and the set of edges $E = \{(v_i, v_j) \mid \text{There is an edge between } v_i \text{ and } v_j\}$. For a given vertex v_i, let $N(v_i)$ be the neighbor vertices of v_i. We have $w(v_i, v_j)$ as the weight of the edge (v_i, v_j). For an undirected graph, $w(v_i, v_j) = w(v_j, v_i)$. Let $w(v_i)$ be the degree of v_i, and we have

$$w(v_i) = \sum_{v_j \in N(v_i)} w(v_j, v_i) = \sum_{v_j \in N(v_i)} w(v_i, v_j). \tag{1}$$

Based on the graph, we can employ various graph-based ranking methods to recommend tags. In this paper, we first introduce two existing methods, namely *most popular tags* and *FolkRank*. Furthermore, we describe FolkDiffusion for social tag suggestions.

3.2   Most Popular Tags

Here we introduce a simple but effective method for tag suggestions. Some notations are given as below, which is identical with [3]. For a user $u \in U$, we denote all his/her tag assignments as $Y_u = Y \cap (\{u\} \times T \times R)$. Accordingly, we have Y_r and Y_t. Based on the same principle, we can define $Y_{u,t} = Y \cap (\{u\} \times \{t\} \times R)$ for $u \in U$ and $t \in T$. We also have $Y_{t,r}$ accordingly. Furthermore, we denote all tags that user $u \in U$ have assigned as $T_u = \{t \in T | \exists r \in R : (u, t, r) \in Y\}$.

There are variants of *most popular tags* as shown in [4], which are usually restricted in different statistical range. For example, *most popular tags of folksonomy* recommends the most popular tags of the whole set of folksonomy. Therefore, it recommends the same set of tags for any given user and resource pair, which suffers from cold-start problems and has no consideration on personalization.

A reasonable variant of *most popular tags* is recommending the tags that globally are most specific to the resource. The method is named as *most popular tags by resource*

$$T(u, r) = \underset{t \in T}{\mathrm{argmax}}^n(|Y_{t,r}|). \tag{2}$$

Since users might have specific preferences for some tags, which should have been used by him/her, thus we can use the *most popular tags by user*. As shown in [4], the performance is poor if we use *most popular tags by user* in isolation. If we mix the *most popular tags by user and resource*, the performance will be much better than each of them in isolation. The simplest way to mix the effect of users and resources on tags is to add the counts and then sort

$$T(u, r) = \underset{t \in T}{\mathrm{argmax}}^n(|Y_{t,r}| \times |Y_{u,t}|). \tag{3}$$

3.3   FolkRank

In FolkRank, two random surfer models are employed on the user-resource-tag tripartite graph. The ranking values of vertices are computed with the following formula

$$PR(v_i) = \lambda \sum_{v_j \in N(v_i)} \frac{w(v_j, v_i)}{w(v_j)} PR(v_j) + (1 - \lambda)p(v_i), \tag{4}$$

where $PR(v_i)$ is the PageRank value and p_{v_i} is the preference to v_i. Suppose we have an adjacent matrix $\mathbf{A}$ to represent the graph $G_{\mathbb{F}}$

$$A(i, j) = \begin{cases} 0 & \text{if } (v_i, v_j) \notin E \\ \frac{w(v_i, v_j)}{w(v_j)} & \text{if } (v_i, v_j) \in E \end{cases}$$

With the matrix, we rewrite the Equation 4 as

$$\mathbf{s} = \lambda \mathbf{A} \mathbf{s} + (1 - \lambda)\mathbf{p}, \tag{5}$$

where $\mathbf{s}$ is the vector of PageRank scores of vertices, and $\mathbf{p}$ is the vector of preferences of vertices.

A straightforward idea of graph-based tag suggestions is to set preferences to the user and resource pair to be suggested for, and then compute ranking values using PageRank with Equation (5). However, as pointed out in [4], this will make it difficult for other vertices than those with high edge degrees to become highly ranked, no matter what the preference values are.

Based on the above analysis, we describe FolkRank as follows. To generate tags for user u and resource r, we have to:

1. Let $\mathbf{s}^{(0)}$ be the stable results of Equation (5) with $\mathbf{p} = \mathbf{1}$, i.e., the vector is composed by 1s.
2. Let $\mathbf{s}^{(1)}$ be the stable results of Equation (5) with $\mathbf{p} = \mathbf{0}$, but $p(u) = \frac{|Y_u|}{|Y|}$ and $p(r) = \frac{|Y_r|}{|Y|}$, where $Y_u = Y \cap (\{u\} \times T \times R)$ and $Y_r = Y \cap (U \times T \times \{r\})$.
3. Compute $\mathbf{s} = \mathbf{s}^{(1)} - \mathbf{s}^{(0)}$.

We rank tags according to their final values in $\mathbf{s}$, where the top-ranked tags are suggested to the given user u and resource r.

3.4   FolkDiffusion

For a graph $G = \{V, E\}$, denote $f_i(t)$ is the heat on vertex v_i at time t, we construct FolkDiffusion as follows. Suppose at time t, each vertex v_i receives an amount of heat, $M(v_i, v_j, t, \Delta t)$, from its neighbor v_j during a period Δt. The received heat is proportional to the time period Δt and the heat difference between v_i and v_j, namely $f_j(t) - f_i(t)$. Based on this, we denote $M(v_i, v_j, t, \Delta t)$ as

$$M(v_i, v_j, t, \Delta t) = \gamma(f_j(t) - f_i(t))\Delta t,$$

where γ is heat diffusion factor, i.e. the thermal conductivity. Therefore, the heat difference at vertex v_i between time $t + \Delta t$ and time t is equal to the sum of the heat that it receives from all its neighbors. This is formulated as

$$f_i(t + \Delta t) - f_i(t) = \sum_{v_j \in N(v_i)} \gamma(f_j(t) - f_i(t))\Delta t. \qquad (6)$$

The process can also be expressed in a matrix form

$$\frac{\mathbf{f}(t + \Delta t) - \mathbf{f}(t)}{\Delta t} = \gamma \mathbf{H} \mathbf{f}(t). \qquad (7)$$

where $\mathbf{f}$ is a vector of heat at vertices at time t, and $\mathbf{H}$ is

$$H(i, j) = \begin{cases} -1 & \text{if } i = j \\ 0 & \text{if } i \neq j \text{ and } (v_i, v_j) \notin E \\ \frac{w(v_i, v_j)}{w(v_j)} & \text{if } i \neq j \text{ and } (v_i, v_j) \in E \end{cases} \qquad (8)$$

If the limit $\Delta t \to 0$, the process will turn into

$$\frac{d}{dt}\mathbf{f}(t) = \gamma \mathbf{H} \mathbf{f}(t). \qquad (9)$$

Solving this differential equation, we have $\mathbf{f}(t) = e^{\gamma t \mathbf{H}} \mathbf{f}(0)$. Here we could extend the $e^{\gamma t \mathbf{H}}$ as

$$e^{\gamma t \mathbf{H}} = \mathbf{I} + \gamma t \mathbf{H} + \frac{\gamma^2 t^2}{2!} \mathbf{H}^2 + \frac{\gamma^3 t^3}{3!} \mathbf{H}^3 + \cdots . \tag{10}$$

The matrix $e^{\gamma t \mathbf{H}}$ is named as the diffusion kernel in the sense that the heat diffusion process continues infinitely from the initial heat diffusion.

γ is an important factor in the diffusion process. If γ is large, the heat will diffuse quickly. If γ is small, the heat will diffuse slowly. When $\gamma \to +\infty$, heat will diffuse immediately, and FolkDiffusion degrades into PageRank.

Like in PageRank, there are random relations between vertices. To capture these relations, we use a uniform random relation between vertices as in PageRank. Let $1 - \lambda$ denote the probability that random surfer happens and λ is the probability of following the edges. Based on the above discussion, we modify FolkDiffusion into

$$\mathbf{f}(t) = e^{\gamma t \mathbf{R}} \mathbf{f}(0), \mathbf{R} = \lambda \mathbf{H} + (1 - \lambda) \frac{1}{N} \mathbf{1}. \tag{11}$$

In application, a computation of $e^{\gamma t \mathbf{R}}$ is time consuming. We usually approximate it to a discrete form

$$\mathbf{f}(t) = (\mathbf{I} + \frac{\gamma}{M} \mathbf{R})^{Mt} \mathbf{f}(0). \tag{12}$$

Without loss of generality, we use one unit time for heat diffusion between vertices and their neighbors. We have

$$\mathbf{f}(1) = (\mathbf{I} + \frac{\gamma}{M} \mathbf{R})^{M} \mathbf{f}(0). \tag{13}$$

We can thus iteratively calculate $(\mathbf{I} + \frac{\gamma}{M} \mathbf{R})^{M} \mathbf{f}(t)$ by applying the operator $(\mathbf{I} + \frac{\gamma}{M} \mathbf{R})$ to $\mathbf{f}(0)$. Therefore, for each iteration, we could diffuse the heat values $\mathbf{s}$ of vertices using the following formulation

$$\mathbf{s} = (1 - \frac{\gamma}{M}) \mathbf{s} + \frac{\gamma}{M} (\lambda \mathbf{A} \mathbf{s} + (1 - \lambda) \frac{1}{N} \mathbf{1}), \tag{14}$$

where M is the number of iterations. As analyzed in [17], for a given threshold ϵ, we can compute to get M such that

$$\|((\mathbf{I} + \frac{\gamma}{M} \mathbf{R})^{M} - e^{\gamma \mathbf{R}}) \mathbf{f}(0)\| < \epsilon, \tag{15}$$

for any $\mathbf{f}(0)$ whose sum is 1.0. Similar to [17], in this paper we set $M = 10$ for FolkDiffusion.

Unlike FolkRank, FolkDiffusion sets the initial values $\mathbf{f}(0)$ for vertices to indicate the preferences. To suggest tags to user u for resource r, we set $\mathbf{f}(0) = \mathbf{0}$, but for $\mathbf{f}_u(0) = 1$ and $\mathbf{f}_r(0) = 1$. After running FolkDiffusion on the tripartite graph, we rank tags according to their heat values and the top-ranked tags are suggested to user u for resource r.

3.5   Dynamic Length of Suggested Tag List

For a social tagging system, it is straightforward to suggest dynamic number of tags based on the characteristics of given users and resources instead of fix number of tags. There are several factors that may influence the number of tags for a post:

1. The global factor, which is the global average number of tags for each post, g;
2. The user factor, which is the average number of tags for each post by this user, u;
3. The resource factor, which is the average number of tags for each post for this resource, r.

Motivated by the idea in [13], with the three factors, we estimate the suggested number of tags sn for a given user and resource using formula $sn = \beta_1 g + \beta_2 u + \beta_3 r$, where β_1, β_2 and β_3 are parameters and can be obtained using linear regression using training data. Note that the dynamic length of suggested tag list is independent to tag suggestion methods. The results in the following experiments are all obtained using dynamic length of tag list.

4   Experiments

In experiments, we compare the results of three graph-based methods: most popular tags, FolkRank and FolkDiffusion. Before demonstrating the experiment results, we introduce the dataset and evaluation metric for experiments.

4.1   Dataset and Evaluation Metric

In experiments we use two datasets provided by BibSonomy. BibSonomy is a social bookmark and publication sharing system that supports users to tag and share URLs and bibtexs. After removing all users, tags, and resources which appear in only one post, the datasets contains 253,615 tag assignments, posted by 1,185 users to 14,443 URLs and 7,946 bibtexs using 13,276 unique tags. Since the user behaviors for tagging URLs and bibtexs are quite different, we perform experiments on bookmark posts (41,268) and bibtex posts (22,852) separately. For convenience, we refer to the two parts as *Bookmark* dataset and *Bibtex* dataset respectively.

   To ensure the statistical validity of experiment results, we carry out experiments using 5-fold cross validation on datasets Bibtex and Bookmark. The performance of tag suggestions is measured using precision, recall and F_1-measure $= \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$.

4.2   Comparing with Other Methods

In Fig. 1, we show the best performance of various graph-based methods on Bibtex and Bookmark datasets when the maximum of suggested tags are set from 1 to 10. In this figure, we also demonstrate the performance of the content-based method kNN, which finds most k similar resources according to their contents and suggest the tags of the k nearest neighbors to the given resource. kNN achieves the best result when $k = 2$. For the method of most popular tags, we use *mpt+resource* to indicate most popular tags by *resource*, and *mpt+mix* to indicate most popular tags by mixing *resource* and *user*. For
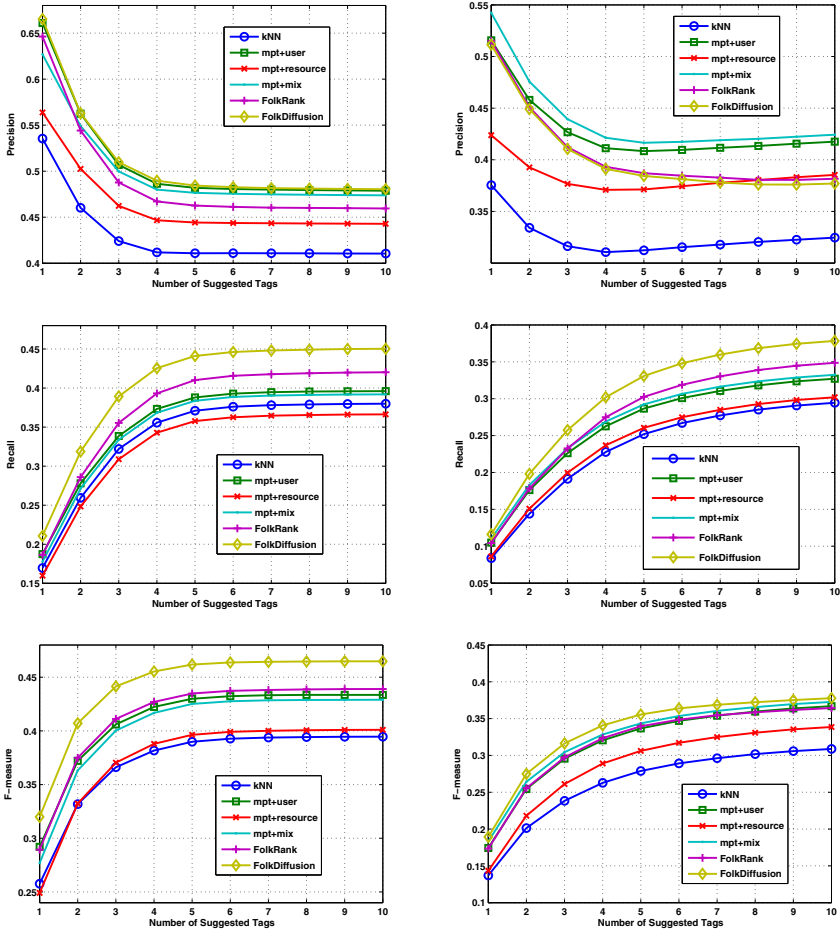
Fig. 1. Performance on the Bibtex and Bookmark datasets

FolkRank, the best result is achieved when damping factor $\lambda = 0.85$ with maximum iterations $max_{it} = 100$. FolkDiffusion obtains the best result when damping factor $\lambda = 0.85$, maximum number of iterations $max_{it} = 10$ and diffusion factor $\gamma = 0.1$.

From the two figures, we have four observations:

1. FolkDiffusion achieves the best F_1-measure on the both datasets. When maximum number of tags is set to 5, FolkDiffusion on Bibtex achieved F_1-measure 0.462 compared to 0.435 of FolkRank. While on Bookmark, FolkDiffusion achieved F_1-measure 0.356 compared to 0.340 of FolkRank and 0.344 of *mpt+mix*.
2. For Bibtex and Bookmark datasets, graph-based methods outperform content-based methods, especially on precision. This is because Bibtex and Bookmark datasets are lack of text information.

3. All graph-based methods perform better on Bibtex than Bookmark. This may be-
cause the resources of Bibtex dataset are bibtexs of scientific articles. Users usually
assign tags to bibtexs mostly based on the themes of the articles, and thus are more
reasonable and easier for predicting.
4. Since we use dynamic length of suggested tag list, the F_1-measure does not drop
when the maximum number of suggested tags are growing.

We also explore the performance of FolkDiffusion on training and test dataset
released in Task-2 of *2009 ECML/PKDD Discovery Challenge* (`http://www.`
`kde.cs.uni-kassel.de/ws/dc09`) [2]. FolkDiffusion achieves precision $=$
0.327, recall $= 0.321$ and $F_1 -$ measure $= 0.3241$, which ranks the fourth place com-
pared to twelve submission results, which is slightly-lower than the performance of
the second and third places. Note that the evaluation results on the released test set are
much worse than those shown in Fig. 1, which is caused by the different distributions of
resources/users/posts between the training set and test set. In all, the evaluation results
verify the effectiveness of FolkDiffusion.

4.3   Influence of Parameter

We further investigate the influence of dynamic length of suggested tag list. If we do not
use the dynamic length and always suggest 5 tags for each given user and resource, the
F_1-measure of FolkDiffusion will drop to only 0.398 on Bibtex and 0.342 on Bookmark.
The influence is also valid to other graph-based methods.

Besides the above analysis, we also investigate the impacts of FolkDiffusion pa-
rameters. In Table 1, we demonstrate the performance of FolkDiffusion on Bibtex and
Bookmark as its parameters change, including the diffusion factor γ and maximum
number of iterations (denoted as *max-it* in tables). Here the damping factor λ is set to
0.85. We find that the performance of FolkDiffusion are not sensitive to its parameters,
which indicates the robustness of FolkDiffusion.

Table 1. Performance of FolkDiffusion on Bibtex (left) and Bookmark (right) dataset. In this
experiment, damping factor λ is set to $\lambda = 0.85$, and the maximum number of suggested tags is
set to 5.

γ	max-it	Precision	Recall	F_1-measure	γ	max-it	Precision	Recall	F_1-measure
1.00	10	0.4758	0.4491	0.4621	1.00	10	0.3834	0.3299	0.3547
0.10	10	0.4844	0.4411	0.4618	0.10	10	0.3878	0.3265	0.3545
0.01	10	0.4772	0.4504	0.4634	0.01	10	0.3844	0.3308	0.3556
0.01	100	0.4772	0.4504	0.4634	0.01	100	0.3844	0.3308	0.3556

5   Conclusion and Future Work

In this paper, we introduce a new graph-based method for tag suggestions, namely
FolkDiffusion, motivated by the idea of heat diffusion. We study the performance of
kNN, most popular tags, FolkRank and FolkDiffusion on real online social tagging
datasets. Experiments show that FolkDiffusion outperforms other graph-based methods

on both Bibtex dataset and Bookmark dataset significantly. This indicates the effectiveness of heat diffusion for modeling user tagging behaviors.

As some recent work indicates, ensembles of various methods will greatly improve the performance of tag suggestions. In future work, we plan to combine FolkDiffusion with other graph-based and content-based methods together to implement a practical social tagging system.

Acknowledgments

References

1. Belkin, M., Niyogi, P.: Laplacian eigenmaps for dimensionality reduction and data representation. Neural computation 15(6), 1373–1396 (2003)
2. Eisterlehner, F., Hotho, A., Jaschke, R.: ECML PKDD Discovery challenge 2009. In: CEUR-WS. org (2009)
3. Jaschke, R., Marinho, L., Hotho, A., Schmidt-Thieme, L., Stumme, G.: Tag recommendations in folksonomies. In: Kok, J.N., Koronacki, J., Lopez de Mantaras, R., Matwin, S., Mladenič, D., Skowron, A. (eds.) PKDD 2007. LNCS (LNAI), vol. 4702, pp. 506–514. Springer, Heidelberg (2007)
4. Jaschke, R., Marinho, L., Hotho, A., Schmidt-Thieme, L., Stumme, G.: Tag recommendations in social bookmarking systems. AI Communications 21(4), 231–247 (2008)
5. Katakis, I., Tsoumakas, G., Vlahavas, I.: Multilabel text classification for automated tag suggestion. In: ECML/PKDD Discovery Challenge 2008, p. 75 (2008)
6. Lafferty, J., Lebanon, G.: Diffusion kernels on statistical manifolds. The Journal of Machine Learning Research 6, 163 (2005)
7. Lipczak, M.: Tag recommendation for folksonomies oriented towards individual users. In: ECML/PKDD Discovery Challenge 2008, p. 84 (2008)
8. Ma, H., Yang, H., King, I., Lyu, M.R.: Learning latent semantic relations from clickthrough data for query suggestion. In: Proceeding of CIKM, pp. 709–718 (2008)
9. Ma, H., Yang, H., Lyu, M.R., King, I.: Mining social networks using heat diffusion processes for marketing candidates selection. In: Proceeding of CIKM, pp. 233–242 (2008)
10. Mishne, G.: Autotag: a collaborative approach to automated tag assignment for weblog posts. In: Proceedings of WWW, pp. 953–954 (2006)
11. Ohkura, T., Kiyota, Y., Nakagawa, H.: Browsing system for weblog articles based on automated folksonomy. In: Proceedings of WWW (2006)
12. Page, L., Brin, S., Motwani, R., Winograd, T.: The PageRank citation ranking: Bringing order to the web. Tech. rep, Stanford University (1998)
13. Rendle, S., Schmidt-Thieme, L.: Factor models for tag recommendation in bibsonomy. In: ECML/PKDD Discovery Challenge (2009)
14. Sood, S., Owsley, S., Hammond, K., Birnbaum, L.: TagAssist: Automatic tag suggestion for blog posts. In: Proceedings of ICWSM, p. 28 (2007)
15. Tatu, M., Srikanth, M., D'Silva, T.: RSDC 2008: Tag recommendations using bookmark content. In: ECML/PKDD Discovery Challenge 2008(2008)
16. Xu, Z., Fu, Y., Mao, J., Su, D.: Towards the semantic web: Collaborative tag suggestions. In: Collaborative Web Tagging Workshop at WWW 2006 (2006)
17. Yang, H., King, I., Lyu, M.R.: DiffusionRank: a possible penicillin for web spamming. In: Proceedings of SIGIR, pp. 431–438 (2007)