A Graph-Coarsening Approach for Tag Recommendation

Manel Hmimida
LIPN-CNRS UMR 7030
University of Paris Nord
99 Av. J.B. Cément 93430
Villetaneuse, FRANCE
manel.hmimida@gmail.com

Rushed Kanawati
LIPN-CNRS UMR 7030
University of Paris Nord
99 Av. J.B. Cément 93430
Villetaneuse, FRANCE
rk@lipn.univ-paris13.fr

ABSTRACT

In this paper we propose a new graph-based tag recommendation approach. The approach is structured into an *offline* step and an *online* one. Offline, the hypergraph depicting the history of tags assignment by users to resources is abstracted. On online, for a given target user and a resource, we first compute the set of recommended *abstract tags* (i.e tag clusters) applying a basic graph-based approach to the abstract graph. A new reduced graph is computed by unfolding the abstract subgraph composed of the set of recommended abstract tags and nodes representing the cluster of users (resp. resources) to which the target user (resp. resource) belongs to. Again the same basic graph-based tag recommendation approach is applied to this new reduced graph in order to compute the final set of tags to recommend. Experiments on real dataset show the effectiveness of the proposed approach.

Keywords

Tag recommendation, Multiplex network, Community detection

1. INTRODUCTION

Social tagging systems, or folksonomies, are popular Web 2.0 tools that allow people to share and organize large sets of resources such as bookmarks, documents, photos, etc. Tag recommendation is a core service in such systems. The goal is to compute the most adequate tag set that a user can apply to annotate a given resource. This helps in controlling the tag vocabulary set, enhancing hence its usefulness for resource access and searching while keeping the annotation process user-centred. This problem has attracted much of interests in the last few years with a variety of different approaches being proposed [5, 8, 9]. Graph-based approaches constitute a major trend in this area. These are attractive approaches since they relay only on mining the induced graph structure of the tagging history making them independent form the type of annotated resources. Actually, the tagging activity history can be represented as a 3-uniform hypergraph where all hyperedges involve three nodes of different types: a user, a resource and a tag. Graph-based tag recommendation approaches include node ranking approaches [5, 6], graph-search based approaches [3], link-prediction approaches [9] and graph-clustering approaches [8]. While graph-based approaches yield interesting results, they often suffer from high execution times due to the large-scale of handled graphs. In this work, we propose a graph-coarsening based approach that can overcome this drawback. The proposed approach is decomposed into two steps: an *offline* step where the folksonomy hypergraph is abstracted by applying a topological clustering approach to the three sets of nodes: users, resources and tags, and an online step during which recommended tags are computed. Upon receiving a query composed of a target user and resource we apply a basic graph-based tag recommendation approach to the abstract graph in order to compute a set of recommended abstract tags. These will be used to construct a new reduced graph, called the *contextual graph* by unfolding the abstract subgraph composed of the set of recommended abstract tags and nodes representing the cluster of users (resp. resources) to which the target user (resp. resource) belongs to. Again the same basic graph-based tag recommendation approach is applied to this new reduced graph in order to compute the final set of tags to recommend. Thus the approach consists in replacing the execution of a standard graph-based tag recommendation approach on a large-scale graph by two executions of the same approach on two reduced graphs. This is expected to drastically reduce the online recommendation computation time. The quality of computed recommendations is also expected to be enhanced since the contextual graph is focused on the query (target user and resource) avoiding taking into account query-irrelevant data. In next section, we give more details about the central step of graph coarsening. First evaluations of the proposed approach are reported and discussed in section 3.

2. GRAPH-COARSENING APPROACH

In order to compute the abstract hypergraph (offline step) we first project the raw hypergraph on each of the three sets: users, tags and resources. The raw hypergraph is approximated by a tripartite graph connecting users, resources and tags. This tripartite graph is first decomposed into three bipartite graphs: Users-Tags, Users-Resources and Resources-tags. Then each of these bipartite graphs is further projected on each of its components. By the end we get three *multiplex*

WWW'16 Companion, April 11–15, 2016, Montréal, Québec, Canada.
ACM 978-1-4503-4144-8/16/04.
http://dx.doi.org/10.1145/2872518.2889415 .

networks defined on the three sets: users, tags and resources. Recall that a multiplex network is a multi-layer network defined over the same set of nodes but each layer contains a different set of edges. We apply a community detection algorithm to each multiplex in order to compute clusters of users, resources and tags. Different approaches for community detection in multiplex networks can be applied[4], including a) *Layer aggregation* (denoted LA) approaches where we first combine all layers and then apply community detection algorithm to the resulting unipartite network, b) *Ensemble clustering* (denoted EC) approaches where we apply a community detection algorithm to each layer then we combine the obtained clusterings, and c) *Multi-layer approaches* that consist in adapting existing algorithms to the multi-layer nature of multiplex networks [4, 7]. The abstract hypergraph is then constructed by replacing each community of each type by a single abstract node.

3. EXPERIMENTS

We have applied the proposed approach to a real dataset extracted from the *Bibsonomy* folksonomy taken from *Het-Rec 2011* [2]. We experimented the approach using *FolkRank* as a *base-line* graph-based tag recommender [5]. Two community detection algorithms are selected, the well known *Louvain* approach [1] and a seed-centric approach developed in our team, the *Licod* algorithm [10]. Both algorithms are used in combination with layer aggregation and ensemble clustering and in their respective generalized versions to multi-layer networks : *GenLouvain*[7] and *MuxLicod* [4]. Another two parameters of the approach are the number of abstract tags to recommend and the number of final tags to recommend. We evaluate the results in terms of both precision and execution time. We have varied the number of abstract tags (denoted $k_{clutser}$) (resp. tags (denoted k_{tag})) to recommend from 1 to 4 (most of resources in the dataset have up to 4 tags). Figure 1 shows the obtained results for $k_{tag} = 3$. The proposed approach yields better results than raw FolkRank with different graph-coarsening approaches but the improvement in terms of precision is rather limited. However, obtained execution times show clearly the advantage of the approach (see table 1), where the execution time drops from 1115 s. to 93 s. when using the Muxlicod algorithm. These executions times are computed for the set of 510 queries composing the test set.

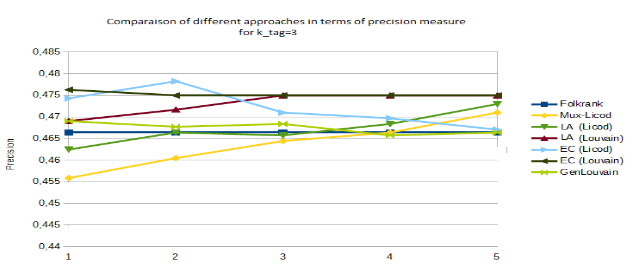

Figure 1: Results in terms of precision for $k_{tag} = 3$

Table 1: Global execution time for all testing queries

Approach	Execution Time (second)
Folkrank	1115
MuxLicod	93
GenLouvain	348
EC(Licod)	154
EC(Louvain)	152
LA(Licod)	850
LA(Louvain)	690

4. CONCLUSION

A graph-coarsening based approach for tag recommendation computation is proposed. The approach yields slightly improved results than raw FolkRank but at much less execution cost. This is a promising result that needs to be confirmed on other datasets and for other basic approaches other than FolkRank. The approach can also be used as a framework for benchamarking and comparing different multiplex community detection algorithms.

5. REFERENCES

[1] V. D. Blondel, J.-l. Guillaume, and E. Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008:P10008, 2008.

[2] I. Cantador, P. Brusilovsky, and T. Kuflik., editors. *2nd workshop on information heterogeneity and fusion in recommender systems (hetrec 2011).*, York, NY, USA, 2011. ACM.

[3] M. Gueye, T. Abdessalem, and H. Naacke:. Strec: An improved graph-based tag recommender. In *RSWeb workshop, RecSys*, 2013.

[4] M. Hmimida and R. Kanawati. Community detection in multiplex networks: A seed-centric approach. *Networks and Heterogeneous Media*, 10(1):71–85, March 2015.

[5] R. Jäschke, L. B. Marinho, A. Hotho, L. Schmidt-Thieme, and G. Stumme. Tag recommendations in folksonomies. In *Knowledge Discovery in Databases: PKDD 2007*, pages 506–514, Warsaw, Poland, 2007.

[6] H.-N. Kim and A. El-Saddik. Personalized pagerank vectors for tag recommendations: inside folkrank. In *ACM conference on Recommender systems*, pages 45–52, 2011.

[7] P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, and J.-P. Onnela. Community structure in time-dependent, multiscale, and multiplex networks. *Science*, 328(5980):876–878, 2010.

[8] S. Papadopoulos, Y. Kompatsiaris, and A. Vakali. A graph-based clustering scheme for identifying related tags in folksonomies. In T. B. Pedersen, M. K. Mohania, and A. M. Tjoa, editors, *DaWak*, volume 6263 of *Lecture Notes in Computer Science*, pages 65–76. Springer, 2010.

[9] M. Pujari and R. Kanawati. Tag recommendation by link prediction based on supervised machine learning. In *Sixth International AAAI Conference on Weblogs and Social Media (ICWSM'2012)*, pages 547–550, Dublin, June 2012.

[10] Z. Yakoubi and R. Kanawati. Licod: Leader-driven approaches for community detection. *Vietnam Journal of Computer Science*, 1(4):241–256, 2014.