Design and Architecture of a Friend Recommender System in the Social Bookmarking Domain

Matteo Manca, Ludovico Boratto, Salvatore Carta
Dipartimento di Matematica e Informatica
Università di Cagliari
Via Ospedale 72, 09124 - Cagliari, Italy
Email: {matteo.manca, ludovico.boratto, salvatore}@unica.it

***Abstract*—Social media systems are becoming more and more popular nowadays. In order to face the overload in the amount of users and content available in these systems, *social recommender systems* have been developed and are largely studied in the literature. A form of social media, known as *social bookmarking system*, allows to share bookmarks in a social network. A user adds as a friend or follows another user and receives updates on the bookmarks added by that user. However, no approach in the literature proposes friend recommender systems in the social bookmarking domain. In this paper, we present an analysis of the state-of-the-art on user recommendation in social environments and of the structure of a social bookmarking system, in order to derive a design and an architecture of a friend recommender system in the social bookmarking domain. This study can be useful for any future research in this area, by highlighting the aspects that characterize this domain and the features that this type of recommender system has to offer.**

***Keywords*—Social Bookmarking; Friend Recommendation; Tagging System.**

I. Introduction

Thanks to the affirmation of social media systems, recent research on recommender system has been focused in this area, by developing a new class of systems known as *social recommender systems* [1], [2]. The information overload, a problem faced by classic recommender systems, in this domain is related to the amount of users and items with whom each user can come across in her/his experience in a social media system. This problem is known as "social interaction overload" [3], [4] and is related to the excessive amount of users and items that each user can interact with. Its consequence is the scarcity of attention, which does not allow to focus on users or items that might be interesting for a user.

So, the role of a social recommender system is to suggest users or items that a user might be interested in. In particular, user recommendation in a social domain aims at suggesting *friends* (i.e., recommendations are built for pairs of users that are likely to be interested in each other's content) or *people to follow* (i.e., recommendations are built for a user, in order to suggest users that might be interesting for her/him) [3].

User recommender systems that operate in the social media domain can be classified into three categories, based on the source of data used to build the recommendations:

1) Systems based on the analysis of social graphs, which explore the set of people connected to the target user in order to produce recommendations. These systems recommend either the closest users in the graph, like friends of friends and followees of followees (the "People you may know" feature offered by Facebook [5] is the most widely known example of this approach), or recommend the users that have the highest probability to be crossed in a random walk of the social graph (the main reference for this type of systems is the "Who to follow" recommendation in Twitter [6]).
2) Systems that analyze the interactions of the users with the content of the system (tags, likes, shares, posts, etc.). In order to exploit the user interests, these systems usually build a user profile by giving a structured form to content, thanks to the use of metrics like TF-IDF (Term Frequency - Inverse Document Frequency). Recommendations are produced by identifying users with similar profiles. An example of this class of systems is presented in [7].
3) Hybrid systems, that consider both the social graph and the interactions of the users with the content (an example is represented by [8]).

An area to which social recommendation can be applied is *social bookmarking systems*. A social bookmarking system allows users to use keywords (*tags*) to describe resources that are of interest for them, helping to organize and share these resources with other users in the social network [9]. The most widely-known examples of social bookmarking systems are Delicious[1], where the bookmarked resources are web pages, and CiteULike[2], where users bookmark academic papers.

To the best of the authors' knowledge, no approach in the literature is able to recommend friends in a social bookmarking system.

In this paper we present a study that proposes the design and the definition of an architecture of a friend recommender system in a social bookmarking system. By analyzing the state-of-the-art on user recommendation in the social domain and how social bookmarking systems work, we design a friend recommender system that operates in this context and present its architecture. Since no other approach in the literature recommends friends in this domain, this study can serve as a reference for any researcher who wants to develop a system in this area.

The scientific contributions coming from this paper are summarized as follows:

[1] http://www.delicious.com
[2] http://www.citeulike.org/

- we analyze the state-of-the-art on user recommendation in social bookmarking systems, in order to highlight the weaknesses of the existing systems and derive the characteristics and features that a friend recommender system that operates in this domain has to offer;
- given the structure of a social bookmarking system and the analysis of the state-of-the-art, we present a design of a friend recommender system;
- we propose a novel architecture of a system to build friend recommendations in a social bookmarking system.

This study can be useful for any future research in this area, by presenting design guidelines and a novel architecture, which can be exploited by any researcher who wants to develop a friend recommender system in the social bookmarking domain.

The rest of the paper is structured as follows: Section II presents the state-of-the-art on user recommendation in social environments; Section III illustrates how a social bookmarking system is structured and how it works; Section IV presents the aspects related to the design of a friend recommender system in a social bookmarking system and presents guidelines, useful in the development of a system; Section V proposes an architecture of the system; Section VI presents conclusions and future work.

II. Related work

In the last years, social bookmarking systems have been studied from different points of view. This section presents related work on user recommendation in this research area. This study of the state-of-the-art will be deepened in Section IV, in order to analyze the aspects that characterize a recommender system that operates in this domain and the weaknesses of the existing approaches.

In [6], Gupta et al. present Twitter's user recommendation service, which is based on shared interests, common connections, and other related factors. The proposed system builds a graph, in which the vertices represent users and the directed edges represent the "follow" relationship; this graph is processed with an open source in-memory graph processing engine called Cassovary. Finally, recommendations are built by means of a user recommendation algorithm for directed graphs, based on SALSA (Stochastic Approach for Link-Structure Analysis). In the next section, we are going to analyze this system, in order to design our proposal.

In [7], Chen et al. describe a people recommender system in an enterprise social network domain. They compare four algorithms, two based on social relationship information and two based on content similarity, and demonstrate that the algorithms that use social information are more capable to find known contacts, while algorithms based on content similarities are better to discover new friends. This approach produces the recommendations by analyzing both the interaction with the content and the interaction with the other users and this characteristic will be analyzed while designing our system.

Guy et al. [10] describe a people recommender system for the IBM Fringe social network. The system uses enterprise information, like org chart relationships, paper and patent co-authorship and project co-membership, which are specific of this social network, so it is hard to take into account this approach when designing our system.

Hannon et al. [8] describe a followee recommender system for Twitter, based on tweets and relationships of their Twitter social graphs. By using this information, they build user profiles and demonstrate how these profiles can be used to produce recommendations. In our proposal, we aim at recommending friendship relationships and not users to follow.

In [11], a recommender system based on collocation (i.e., the position of the user) is presented. It uses short-range technologies of mobile phones, in order to infer the collocation and other correlated information that are the base for the recommendations. In our domain we do not have such a type of information, so we cannot compare with this algorithm.

Zhou et al. [12] propose a framework for users' interest modeling and interest-based user recommendation (it suggests people to follow and not friends), which was tested on the Yahoo! Delicious dataset. Recommendations are produced by analyzing the network and fans properties. Differently from this framework, our proposal aims at producing friend recommendations.

In [13], a study about what cues in a user's profile, behavior, and network might be most effective in recommending people, is presented. So, this approach analyzes both the interaction with the content and the interaction with the other users and this characteristic will be analyzed while designing our system.

Liben-Nowell and Kleinberg [14] studied the user recommendation problem as a link prediction problem. They develop several approaches, based on metrics that analyze the proximity of nodes in a social network, to infer the probability of new connections among users. Experiments show that the network topology is a good tool to predict future interactions.

In [15], Arru et al. propose a user recommender system for Twitter, based on signal processing techniques. The considered approach defines a pattern-based similarity function among users and makes use of a time dimension in the representation of the users profile. Our system is different, because we aim at suggesting friends while on Twitter there is no notion of "friend" but it works with "people to follow".

III. Social Bookmarking Systems

This section presents how a social bookmarking system is structured and how it works. This definition is based on the ones previously given in the literature (in particular we refer to [9], [16], [17]).

A social bookmarking system is composed by:

- a set of *users*;
- a set of *resources*. These resources characterize the type of social bookmarking system and, as mentioned in the introduction, they might be of different types (e.g., web pages);
- a set of *tags*, which are the keywords used to describe the resources;

- a set of *bookmarks*, which are represented as triplets $(user, resource, tag)$; these triplets are known either as *tag assignments*, or as *tag applications*;
- a set of *connections* among users, which are represented as couples $(user, user)$. Depending on the type of connection among two users, a couple might be ordered (i.e., users are connected by a *follow* relation), or not (i.e., users are *friends* and mutually follow each other). These connections form a graph, known either as *social graph* or *interest graph*.

Once a user decides to bookmark a resource by adding tags to it, these bookmarks are shown to the users who are friends with or follow this user.

Social bookmarking systems also offer privacy options, which allow to keep a bookmark private, or to share it only with a limited amount of users.

Features that allow to explore the tags and to facilitate the management of the bookmarks, like their export from browsers [18] and the possibility to add a bookmark to the profile by email, are often offered.

IV. Designing a Friend Recommender System

The first objective of our proposal is to design a friend recommender system in a social bookmarking system. This section presents an analysis of the aspects that characterize both the state-of-the-art and social bookmarking systems, according to what was presented in the previous sections.

A. Analysis

In our analysis, we considered the following aspects:

(a) In [6], authors highlight that Twitter is an "interest graph", rather than a "social graph". A problem highlighted by the authors is that the analysis of such a graph suffers from scalability issues and, in order to contain the complexity of the recommender system, no user profile information could be used to produce the recommendations. The definition of interest graph can also be extended to social bookmarking systems, since a user can add as a friend or follow another user, in order to receive her/his newly added bookmarks.
(b) Social media systems grow rapidly. This means that both the amount of content added to a social media system and the user population increase at a fast rate. A recommender system that operates in this context needs to build accurate user profiles, which are up-to-date with the constantly evolving preferences of the users.
(c) Resources usually have an unstructured form so, when building a content-based recommender systems, they are given a structured form, by introducing a *Content Analyzer* in the system [19].
(d) In the architecture of a content-based system, a *Feedback* component, which allows to update the user profile according to the recommended items that the user liked or did not like, is usually present [19].
(e) As [12] highlights, the tagging activity of the users reflects their interests. Therefore, the tags used by a user can be considered as an important source of information to exploit her/his interests.

Taking into account all these aspects, we drew the following conclusions.

Regarding point (a), in order to avoid the limitations related to the graph analysis in this domain, we aim at designing a system that only analyzes the content of the users (i.e., the tagged resources). So, we are going to design a system that belongs to the second class presented in the Introduction, i.e., the one that analyzes the interactions of the users with the content of the system.

Regarding points (b) and (c), in order to efficiently and quickly update user profiles, we decided to exploit the set of resources used by each user and the tags used to classify those resources, without using a *Content Analyzer* component, but analyzing only the behavior of the users on the system.

Regarding point (d), since the system we are designing deals with friend recommendations and we do not consider the interest graph, the feedback of a user has no impact in her/his profile. On the contrary, when items are recommended in a content-based system, the feedbacks contains information about the preferences of the users, which help updating the user profiles.

Regarding point (e), we embraced the theory that user interests are reflected by the tagging activity and extended it, by following the intuition that users with similar interests use similar tags and the same resources.

B. Design guidelines

Starting from the previous analysis, here we recap what we believe are the features that a friend recommender system in the social bookmarking domain has to offer:

1) a friend recommender system in a social bookmarking system has to consider the tagged resources bookmarked by the users. Using only graph analysis to build the recommendations presents limitations, and the production of recommendations by analyzing both the content the users interacts with and the interest graph would increase the complexity of the system (this might lead to the learning of user profiles that are not up-to-date with the current interests of the users);
2) the algorithms and metrics used by a system should be quickly computed, in order to keep the user profiles up-to-date. Therefore, we believe that a friend recommender system should mine *user behavior* (i.e., the interaction of the users with the content), more than the content itself. In fact, the introduction of a *Content Analyzer*, in order to give a structured form to the resources, would significantly increase the complexity of the system. In other words, it is harder to make an analysis of the content of each resource tagged by a user, instead of considering only the fact that a user is interested in that resource. Since social bookmarking systems grow at a fast rate, content analysis would lead to have outdated profiles and this component is discarded by our design and architecture;
3) in order to reduce the complexity of the system, and given the type of recommendations produced, the typical *Feedback* component of a Content-Based recommender system is removed when designing such a type of system.

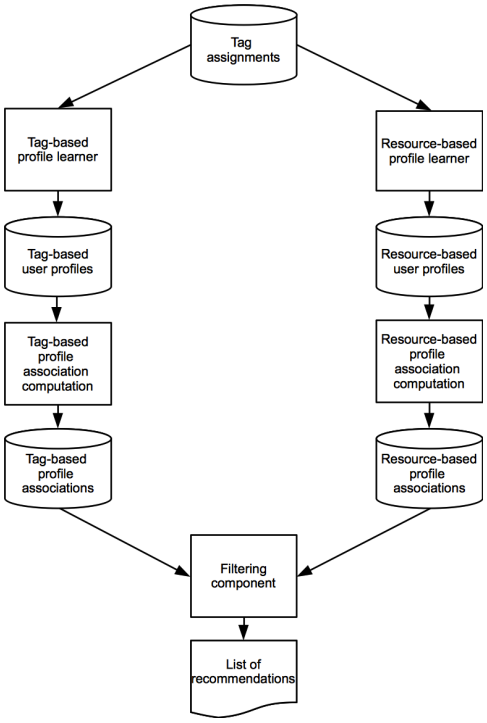

Fig. 1: Architecture of the friend recommender system

This choice was made since the accepted or rejected friends do not update the user profiles, which are built considering the tag assignments of the users;

4) in order to capture the interaction of the users on multiple levels and improve the capability to accurately recommend friends, a system has be able to exploit all the sources of information coming from the tag assignments. Therefore, a friend recommender system has to analyze both the tags used by a user and the resources she/he bookmarked.

V. Architecture

The objective of this section is to present a novel architecture for friend recommendation in a social bookmarking system, by considering the design guidelines presented in the previous section. Figure 1 illustrates the architecture.

While designing the system, in the first point of the guidelines we highlighted that we would only analyze the content of the system (i.e., the tag assignments). Therefore, the architecture does not have components that analyze the interest graph, which contains the connections among users.

As the figure shows, our architecture presents two profile learners (i.e., the *Tag-based profile learner* and the *Resource-based profile learner*), following the considerations done on the fourth point of the design guidelines, which suggested to consider both the tags and the resources available in the bookmarks.

The components that form the proposed architecture are now described in detail.

- **Tag-based profile learner.** Taken as input the *Tag assignments* available for each user, this component builds a user profile, by considering the tags used by a user. Since in the design guidelines we highlighted the need to build profiles quickly, in order for them to be updated, this component might build profiles as binary vectors of the tags considered by users, or by considering the frequencies of each tag used by a user. The output produced by this component is a *Tag-based user profile*.
- **Resource-based profile learner.** Given the *Tag assignments* available for each user, this component builds a second user profile, by analyzing the resources bookmarked by a user. Also this profile might be built as a vector, similarly to the possible implementations of the tag-based component. The output produced by this component is a *Resource-based user profile*.
- **Tag-based profile association computation.** Given the *Tag-based user profiles* previously computed, this step estimates the association among each couple of tag-based user profiles, in order to derive how similar two users are. The output produced by this component is a *Tag-based profile association*, i.e., a number that quantifies this similarity among each couple of users.
- **Resource-based profile association computation.** Given the *Resource-based user profiles* previously computed, this step estimates the association among each couple of resource-based user profiles, in order to derive how similar two users are. The output produced by this component is a *Resource-based profile association*, i.e., a number that quantifies this similarity among each couple of users.
- **Filtering component.** Given the *Tag-based profile associations* and *Resource-based profile associations* previously built, this component filters the similarities, in order to select the most similar users to recommend to the target user. This component ranks the associations and filters them. For example, a threshold value might be used, in order to select only the users with high associations with the target user. The output of this component is a ranked *List of recommendations*, which contains the users to recommend to the target user.

VI. Conclusions and Future Work

This paper illustrated a study related to the design and the architecture of a friend recommender system in the social bookmarking domain. Since no other approach in the literature recommends friends in this area, we analyzed the existing state-of-the-art works that recommend users in social domain and illustrated the structure of a social bookmarking system. This led to the design of a system that recommends friends in this context. After giving the design guidelines, the architecture of the system was presented.

We are currently implementing a friend recommender system that is based on the design and the architecture presented in this paper.

As a future work we will test the system's capability to accurately and efficiently recommend friends.

ACKNOWLEDGMENT

This work is partially funded by Regione Sardegna under project SocialGlue, through PIA - Pacchetti Integrati di Agevolazione "Industria Artigianato e Servizi" (annualità 2010).

REFERENCES

[1] Francesco Ricci, Lior Rokach, and Bracha Shapira, "Introduction to recommender systems handbook," in *Recommender Systems Handbook*, Francesco Ricci, Lior Rokach, Bracha Shapira, and Paul B. Kantor, Eds., pp. 1–35. Springer, 2011.

[2] Patricia Victor, Martine De Cock, and Chris Cornelis, "Trust and recommendations," in *Recommender Systems Handbook*, Francesco Ricci, Lior Rokach, Bracha Shapira, and Paul B. Kantor, Eds., pp. 645–675. Springer, 2011.

[3] Ido Guy, Li Chen, and Michelle X. Zhou, "Introduction to the special section on social recommender systems," *ACM TIST*, vol. 4, no. 1, pp. 7, 2013.

[4] Herbert A. Simon, "Designing organizations for an information rich world," in *Computers, communications, and the public interest*, Martin Greenberger, Ed., pp. 37–72. Johns Hopkins Press, Baltimore, 1971.

[5] Florin Ratiu, "Facebook: People you may know," May 2008.

[6] Pankaj Gupta, Ashish Goel, Jimmy Lin, Aneesh Sharma, Dong Wang, and Reza Zadeh, "Wtf: the who to follow service at twitter," in *22nd International World Wide Web Conference, WWW '13, Rio de Janeiro, Brazil, May 13-17, 2013*, Daniel Schwabe, Virgílio A. F. Almeida, Hartmut Glaser, Ricardo A. Baeza-Yates, and Sue B. Moon, Eds. 2013, pp. 505–514, International World Wide Web Conferences Steering Committee / ACM.

[7] Jilin Chen, Werner Geyer, Casey Dugan, Michael J. Muller, and Ido Guy, "Make new friends, but keep the old: recommending people on social networking sites," in *Proceedings of the 27th International Conference on Human Factors in Computing Systems, CHI 2009, Boston, MA, USA, April 4-9, 2009*, Dan R. Olsen Jr., Richard B. Arthur, Ken Hinckley, Meredith Ringel Morris, Scott E. Hudson, and Saul Greenberg, Eds. 2009, pp. 201–210, ACM.

[8] John Hannon, Mike Bennett, and Barry Smyth, "Recommending twitter users to follow using content and collaborative filtering approaches," in *Proceedings of the 2010 ACM Conference on Recommender Systems, RecSys 2010, Barcelona, Spain, September 26-30, 2010*, Xavier Amatriain, Marc Torrens, Paul Resnick, and Markus Zanker, Eds. 2010, pp. 199–206, ACM.

[9] Umer Farooq, Thomas George Kannampallil, Yang Song, Craig H. Ganoe, John M. Carroll, and C. Lee Giles, "Evaluating tagging behavior in social bookmarking systems: metrics and design heuristics," in *Proceedings of the 2007 International ACM SIGGROUP Conference on Supporting Group Work, GROUP 2007, Sanibel Island, Florida, USA, November 4-7, 2007*, Tom Gross and Kori Inkpen, Eds. 2007, pp. 351–360, ACM.

[10] Ido Guy, Inbal Ronen, and Eric Wilcox, "Do you know?: recommending people to invite into your social network," in *Proceedings of the 2009 International Conference on Intelligent User Interfaces, February 8-11, 2009, Sanibel Island, Florida, USA*, Cristina Conati, Mathias Bauer, Nuria Oliver, and Daniel S. Weld, Eds. 2009, pp. 77–86, ACM.

[11] Daniele Quercia and Licia Capra, "Friendsensing: recommending friends using mobile phones," in *Proceedings of the 2009 ACM Conference on Recommender Systems, RecSys 2009, New York, NY, USA, October 23-25, 2009*, Lawrence D. Bergman, Alexander Tuzhilin, Robin D. Burke, Alexander Felfernig, and Lars Schmidt-Thieme, Eds. 2009, pp. 273–276, ACM.

[12] Tom Chao Zhou, Hao Ma, Michael R. Lyu, and Irwin King, "Userrec: A user recommendation framework in social tagging systems," in *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2010, Atlanta, Georgia, USA, July 11-15, 2010*, Maria Fox and David Poole, Eds. 2010, AAAI Press.

[13] Michael J. Brzozowski and Daniel M. Romero, "Who should i follow? recommending people in directed social networks," in *Proceedings of the Fifth International Conference on Weblogs and Social Media, Barcelona, Catalonia, Spain, July 17-21, 2011*, Lada A. Adamic, Ricardo A. Baeza-Yates, and Scott Counts, Eds. 2011, The AAAI Press.

[14] David Liben-Nowell and Jon M. Kleinberg, "The link prediction problem for social networks," in *Proceedings of the 2003 ACM CIKM International Conference on Information and Knowledge Management, New Orleans, Louisiana, USA, November 2-8, 2003*. 2003, pp. 556–559, ACM.

[15] Giuliano Arru, Davide Feltoni Gurini, Fabio Gasparetti, Alessandro Micarelli, and Giuseppe Sansonetti, "Signal-based user recommendation on twitter," in *22nd International World Wide Web Conference, WWW '13, Rio de Janeiro, Brazil, May 13-17, 2013, Companion Volume*, Leslie Carr, Alberto H. F. Laender, Bernadette Farias Lóscio, Irwin King, Marcus Fontoura, Denny Vrandecic, Lora Aroyo, José Palazzo M. de Oliveira, Fernanda Lima, and Erik Wilde, Eds. 2013, pp. 941–944, International World Wide Web Conferences Steering Committee / ACM.

[16] Shilad Sen, Shyong K. Lam, Al Mamunur Rashid, Dan Cosley, Dan Frankowski, Jeremy Osterhouse, F. Maxwell Harper, and John Riedl, "tagging, communities, vocabulary, evolution," in *Proceedings of the 2006 ACM Conference on Computer Supported Cooperative Work, CSCW 2006, Banff, Alberta, Canada, November 4-8, 2006*, Pamela J. Hinds and David Martin, Eds. 2006, pp. 181–190, ACM.

[17] Andreas Hotho, Robert Jäschke, Christoph Schmitz, and Gerd Stumme, "Bibsonomy: A social bookmark and publication sharing system," in *Proceedings of the First Conceptual Structures Tool Interoperability Workshop at the 14th International Conference on Conceptual Structures*, 2006, pp. 87–102.

[18] Ben Lund, Tony Hammond, Timo Hannay, and Martin Flack, "Social bookmarking tools (ii): A case study - connotea," *D-Lib Magazine*, vol. 11, no. 4, 2005.

[19] Pasquale Lops, Marco de Gemmis, and Giovanni Semeraro, "Content-based recommender systems: State of the art and trends," in *Recommender Systems Handbook*, Francesco Ricci, Lior Rokach, Bracha Shapira, and Paul B. Kantor, Eds., pp. 73–105. Springer, 2011.