Expert Systems with Applications 41 (2014) 457–469

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Summarizing dynamic Social Tagging Systems

Hans-Henning Gabriel [a,1], Myra Spiliopoulou [b], Alexandros Nanopoulos [c,*]

[a] Datameer, USA
[b] Faculty of Computer Science, Otto-von-Guericke University, Magdeburg, Germany
[c] University of Eichstätt-Ingolstadt, Germany

ARTICLE INFO

Keywords:
Dynamic Social Tagging Systems
Social network summarization
Folksonomy
Clustering
Tensor-based clustering

ABSTRACT

Social tagging is a popular method that allows users of social networks to share annotation in the form of keywords, called tags, assigned to resources. Social tagging addresses information overload by easing the task of locating interesting entities in a social network. Nevertheless, users can still be overwhelmed by too many tags posted at each moment. A process is needed that offers an accurate overview of the *representative* entities and their relationships with each other, while dealing with the dynamics of social tagging and of tags' semantics. We propose a method for the automated summarization of an evolving multi-modal social network, focusing on the entities that *stay representative over time* for some subnetwork in the social tagging system. We report on experiments with real data from the Bibsonomy social tagging system, where we compare our dynamic approach with a static one.

1. Introduction

Social Tagging Systems (STSs) become of increasing importance, because they allow users to assign keywords (called 'tags') to any type of content, *and* to share them with other users. People use tags to characterize content semantically. By overtaking tags posted by other people, they also overtake the semantics associated with that content and use the tags to find the same content again. STSs offer a variety of services, such as bookmaking and sharing arbitrary content in del.icio.us, tagging books in LibraryThing, scientific articles in Bibsonomy or news stories in Digg and SlashDot. Yelp and CitySearch allow their users to provide business and product reviews. In Youtube and last.fm users tag multimedia content. Tags are also used to express the categorical decisions from users of search engines such as Yahoo! or Answers, by tagging the answers as helpful or not.

Social tagging find many applications (Gupta et al., 2011): It is particularly useful for both personalized and enterprise search – the latter application is called enterprise bookmarking. The reason is that social tags act as multi-faceted descriptors of content and, thus, improve its findability. Especially in the context of enterprise bookmarking, social tags affect the way people and organizations share information through intranets and document management systems. This offers improved decision-making support both to groups of decision makers within an enterprise and across the enterprise, allowing them to use tags to enhance the findability of content without waiting the (usually slow) process of its formal categorization and cataloging. Other applications in the business domain include using tags as source of information when ranking web sites, since the number of social tags assigned to a site can measure its popularity. Moreover, social tags provide faster and more thorough indexing of web sites, because they allow the discovery of sites that have not been yet linked by others or crawled by search engines. Finally, social tags can support the classication of fast changing information, such as blog entries, whereas they can become valuable for the task of social interest discovery, i.e., finding users' interests and their communities (see Gupta et al. (2011) for a detailed view of applications).

STSs accommodate a huge number of resources associated with semantically disperse tags and their content is updated at fast pace as new resources are uploaded and tags are attached to existing resources. As a result, the users of STSs can become overwhelmed and their experience with respect to all aforementioned applications may be compromised. To assist users in acquiring an overview of an STS as it evolves, we propose a method that maps the evolving network into a static summary. This summary consists of a small set of entities that are *representative* of some part of the STS over time, and a small set of *neighbors* frequently associated to each of these entities over time. Our approach extends the work of Gabriel et al. (2010, 2011) which was designed for static STSs.

Which entities are representative of an evolving STS? Intuitively, such entities are influential users, popular tags or frequently

* Corresponding author. Address: Katholische Universitt Eichsttt-Ingolstadt, Auf der Schanz 49, 85049 Ingolstadt, Germany. Tel.: +49 8419371858; fax: +49 8419372871.

E-mail addresses: hanshenning.gabriel@gmail.com (H.-H. Gabriel), myra@iti.cs.uni-magdeburg.de (M. Spiliopoulou), alexandros.nanopoulos@ku.de (A. Nanopoulos).

[1] Work done while with the Faculty of Computer Science, Otto-von-Guericke University, Magdeburg, Germany.

0957-4174/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2013.07.071

accessed resources, i.e., entities of any types (called modes) that can be found in an STS. They should persistently *represent* some part of the multi-modal STS as it evolves over time. Not surprisingly, methods that monitor communities in an evolving social network, e.g., Sun et al. (2006) and Lin et al. (2008), identify entities that have a central role in these evolving communities. Albeit the focus of these studies is on monitoring the communities, one would expect that the additional step of summarizing the evolving STS into a smaller, comprehensible and comprehensive graph would naturally follow the findings of these studies. However, the summarization and visualization of a summary graph over a large static STS, as proposed, e.g., in Gabriel et al. (2011), does not transfer naturally into the dynamic scenario. Should the summary graph of a dynamic STS be a *sequence of graphs*, as in the ContexTour approach of Lin et al. (2010), or should it be a *single graph* as the TimeFall model of Ferlez et al. (2008) or the Topic Table of Gohr et al. (2010)?.

In this study, we opt for mapping an evolving social tagging system into a *single* summary rather than a sequence of summaries. We anticipate that the single summary allows the casual observer to assign the observed entities into conceptual categories and gain insights to their meaning. This corresponds to the concept of *sensemaking* of Social Tagging Systems, elaborated in Golder and Huberman (2006). According to Weick et al. (2005): 'sensemaking is a process in which information is categorized and labeled and, critically, through which meaning emerges' and proceed in elaborating on the role of tags in sensemaking.

Deriving a single summary over an evolving STS would be fairly straightforward, if all entities were documents: one could use the underpinnings of topic evolution to identify the dominant topics over time, and depict their change as done in Ferlez et al. (2008) and Gohr et al. (2010). However, an STS is multi-modal: a user identifier like '123' or 'immunoassay' does not provide much insight to what this user stands for, no matter how influential the user is; the same holds for a frequently used tag like 'super'. Hence, we need a summarization mechanism that depicts *representative entities* for the evolving STS *and* associates them with *semantics* that would allow an observer to make sense out of them.

In this study, we build upon the summarization of the *CrossSense* algorithm, proposed in Gabriel et al. (2010) for a static STS. We extend *CrossSense* for the case of evolving STSs. The extended method, called *DynamicCrossSense*, identifies entities (users, resources, tags) of the multi-modal STS that *stay representative* of some part of the network over time; these are the *pivot entities* in the STS summary. To complete the summary, *DynamicCrossSense* selects for each pivot a number of *neighbor entities* that are frequently associated with the pivot (e.g., resources often annotated with a pivot tag by different users); these neighbors constitute the *world* of a pivot. Visualization of the summary is beyond the scope of this paper. However, the visualization algorithm we proposed in Gabriel et al. (2011) can be used to represent the summary as a graph of stars, where the center of each star is a pivot, and the links connecting a neighbor to a star center are weighted with some scoring function, e.g., the frequency of co-occurrence between neighbor and pivot.

The paper is organized as follows. In the next section we discuss related work. In Section 3 we formalize the core concepts of our approach, extending the formalism of *CrossSense* (Gabriel et al., 2010) for the case of dynamic STSs. In Section 4 we extend the formalism to deal with the volatility and dynamics of an STS. We use this extended formalism for our *DynamicCrossSense* algorithm, presented in Section 5. The evaluation of this algorithm on real data from the Bibsonomy STS is reported in Section 6. The last section concludes our study and provides an outlook for further research.

2. Related work

The method we propose delivers a summary of an evolving STS system. This summary consists of a selection of entities that *stay representative* of part of the network for some time, and a selection of their close *neighbors*. Related work comes from the area of community evolution monitoring; relevant are also methods that visualize representative entities.

2.1. Monitoring community evolution

Three examples of entity types in a dynamic social network are: users active in a social tagging environment, resources contributed by these users, and tags assigned to resources by other users. In the previous years, there is a growing body of research on capturing the dynamics of social networks, albeit earlier methods focus only on one type of entity in the network. The algorithms TimeFall (Ferlez et al., 2008) and CoDym (Falkowski et al., 2006) provide visualization aids for the overview of an evolving social network, but focus on one type of entity in the social network.

Chakrabarti et al. (2006) and Chi et al. (2007) propose the concept of temporal smoothness, tracing communities that evolve smoothly over time. Ahmed and Xing incorporate temporal smoothness as a parameter in the learner (Ahmed and Xing, 2009), while Zhou et al. propose a parameter-free learner that monitors evolving communities under the assumption of temporal smoothness (Zhou et al., 2010). All these methods focus on understanding how the network evolves, not in providing an overview of it.

A social tagging system encompasses three entity types: users, resources and tags. The monitoring of evolving communities that encompass all three entity types is typically pursued with tensor-based clustering methods. Recent advances include METAFAC (Lin et al., 2009), FacetNet (Lin et al., 2008) and GraphScope (Sun et al., 2007), upon which METAFAC builds. As with the previous methods, these algorithms focus on monitoring the evolving communities, rather than providing an overview of their evolution.

2.2. Depicting and visualizing representative entities in an evolving network

Intuitively, an algorithm that ranks entities in an STS, such as variants of PageRank (Page et al., 1998), e.g., the FolkRank algorithm (Hotho et al., 2006), are in general related to the problem of detecting representative entities. Such ranking algorithms, based on the underlying (hyper-) graph, can choose its top-ranked nodes and build clusters around them. However, such an approach is not appropriate for an evolving network. The reason is that ranking algorithm, being evolutionary by nature, converges to a state that depends on *all* data it has seen. New data (e.g., new social interactions) can be added, but they serve in refining the learned model. In contrast, if some old data are forgotten, the algorithm must be restarted from the earliest data point onwards. This procedure is not practical, and it will favor old entities anyway.

In Lin et al. (2010), Lin et al. propose ContexTour, a visualization mechanism that provides an overview of a dynamic social network from different perspectives, such as the contents of the network and the interactions among its members. ContexTour builds upon the tensor-based algorithm METAFAC in Lin et al. (2009), which derives communities over an evolving multi-faceted framework - by modeling resources, users, tags and their relationships in multiple tensors. Similarly to our approach, ContexTour focusses on identifying important entities in the social network, but it first identifies important *clusters*. We rather believe that clusters found under different perspectives should not be weighted but rather contribute to the importance of individual entities. As a result, the summary