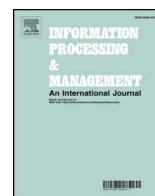




Contents lists available at ScienceDirect

# Information Processing and Management

journal homepage: [www.elsevier.com/locate/infoproman](http://www.elsevier.com/locate/infoproman)

## Mining user interests over active topics on social networks

Fattane Zarrinkalam<sup>a,b</sup>, Mohsen Kahani<sup>\*,b</sup>, Ebrahim Bagheri<sup>a</sup><sup>a</sup> Laboratory for Systems, Software and Semantics (LS<sup>3</sup>), Ryerson University, Canada<sup>b</sup> Department of Computer Engineering, Ferdowsi University of Mashhad, Iran

### ARTICLE INFO

#### Keywords:

User interest detection  
 Social network analysis  
 Homophily  
 Topic relatedness  
 Semantic annotation

### ABSTRACT

Inferring users' interests from their activities on social networks has been an emerging research topic in the recent years. Most existing approaches heavily rely on the explicit contributions (posts) of a user and overlook users' *implicit interests*, i.e., those potential user interests that the user did not explicitly mention but might have interest in. Given a set of active topics present in a social network in a specified time interval, our goal is to build an interest profile for a user over these topics by considering both explicit and implicit interests of the user. The reason for this is that the interests of free-riders and cold start users who constitute a large majority of social network users, cannot be directly identified from their explicit contributions to the social network. Specifically, to infer users' implicit interests, we propose a graph-based link prediction schema that operates over a representation model consisting of three types of information: user explicit contributions to topics, relationships between users, and the relatedness between topics. Through extensive experiments on different variants of our representation model and considering both homogeneous and heterogeneous link prediction, we investigate how topic relatedness and users' homophily relation impact the quality of inferring users' implicit interests. Comparison with state-of-the-art baselines on a real-world Twitter dataset demonstrates the effectiveness of our model in inferring users' interests in terms of perplexity and in the context of retweet prediction application. Moreover, we further show that the impact of our work is especially meaningful when considered in case of free-riders and cold start users.

### 1. Introduction

With the emergence and growing popularity of online social networks such as Twitter, many users extensively use social posts to express their feelings and views about a wide variety of social events/topics as they happen in real time. This has made social networks as a viable source of information about users' interests with regards to the current active topics/events (Abel, Gao, Houben, & Tao, 2011). For instance, when looking at Twitter data during November 2010, the rivalry between the two English Premier League football teams, Spurs and Arsenal is a topic that has attracted a lot of discussion and interest. The development of techniques that can automatically detect such topics and model users' interests towards them from user activities in social networks has become an emerging research area in the recent years, which has the potential to improve the quality of applications that work on a user modeling basis, such as filtering twitter streams (Kapanipathi, Orlandi, Sheth, & Passant, 2011), news recommendation (Abel et al., 2011; Meguebli, Kacimi, Doan, & Popineau, 2017), retweet prediction (Feng & Wang, 2013) and hashtag recommendation (Li, Jiang, Liu, Qiu, & Sun, 2017), among others.

Most existing approaches for detecting users' interests rely heavily on the explicit contributions (posts) of a user (Abel et al., 2011;

\* Corresponding author.

E-mail address: [kahani@um.ac.ir](mailto:kahani@um.ac.ir) (M. Kahani).

Yang, Sun, Zhang, & Mei, 2012). In other words, to detect a user's interests, these approaches predominantly consider the content that the user has posted, shared, viewed or favorited on her social profile. However, they struggle to identify a user's interests if the user has not explicitly mentioned them. For example, consider the following tweets posted by a user, who we call 'Mary':

- “The opportunity to go top of the Premier League will give Arsenal an extra incentive to beat Spurs, according to Wenger <http://bit.ly/chgPjO>”
- “Arsenal won't win with Wenger's policy. Spurs continue to exceed expectations”

Based on the keywords explicitly mentioned in Mary's tweets, one could easily infer that she is interested in the rivalry between Spurs and Arsenal. The interests that are directly observable in a user's tweets are referred to as *explicit interests*. Expanding on this example, another topic that emerged in 2010 was Prince William's engagement. Looking at Mary's posts she never referred to this topic in her tweet stream. However, it is possible that Mary is British and is interested in both football and the British Royal family; although, she never explicitly tweeted about the latter. If that is in fact the case, then Mary's interest profile would need to include an interest pertaining to the British Royal family. We refer to these concealed user interests as *implicit interests*, i.e., topics that the user did not explicitly engage with but might have interest in. The identification of implicit interests has received less attention in the literature but is of significant importance for the purpose of accurate user modeling especially for users who are not active or only tend to be passive consumers of content and consequently their available textual content is sparse and does not reveal sufficient clues about their interests (Spasojevic, Yan, Rao, & Bhattacharyya, 2014).

The main objective of our work in this paper is to build a user interest profile by considering both explicit and implicit interests of the user in a given time interval from Twitter. Based on the homophily principle (McPherson, Smith-Lovin, & Cook, 2001), users tend to interact with users with common interests or preferences. Therefore, interactions between users can be considered to be an important clue for inferring their interests (Wen & Lin, 2010). In addition, relatedness between topics is another important aspect that can uncover implicit interests of users (Bhattacharya, Zafar, Ganguly, Ghosh, & Gummadi, 2014; Shen, Wang, Luo, & Wang, 2013). In this work, we combine these two factors into a unified heterogeneous representation model to consider them simultaneously.

More succinctly, the key contributions of our work are as follows:

- We model users' interests based on their inclination towards the active topics on Twitter. The literature is abundant with techniques that automatically detect active topics from social networks (Alvarez-Melis & Saveski, 2016). In our work, we develop techniques that determine, whether or not, and to what extent, a user is interested in these active topics on Twitter by considering both explicit and implicit interests of the user.
- We propose a graph-based link prediction framework to determine the so-called *implicit interests* of a given user. Our work considers a heterogeneous graph that includes three types of information: *i*) user explicit contributions to active topics, *ii*) the relatedness between those active topics, and *iii*) relationship between users to incorporate theory of homophily. To the best of our knowledge, the proposed framework is among the first to provide such a holistic approach for identifying users implicit interests.
- Another important contribution of our work is the investigation of whether it is possible to identify user interests for those users who are not highly active on the social network either because they are cold start users or free riders. This is especially important when considering the fact that studies have reported that most users on social networks can exhibit cold start or free riding behavior (Romero, Galuba, Asur, & Huberman, 2011). As such, methods that rely on explicit user information for determining user interests might not work well under these circumstances. In our experiments, we will additionally show that the impact of our work is especially meaningful when considered in such context.

The rest of the paper is organized as follows: Section 2 reviews the related work. The proposed approach is introduced in Sections 3. Section 4 is dedicated to the experiments and evaluation of the proposed model, which is followed in Section 5 by a discussion about our motivations and research findings. Finally, Section 6 sheds light on future work and concludes the paper.

## 2. Related work

Our work in this paper first extracts active topics from Twitter and then determines a given user's inclination towards these active topics. We assume that an existing state of the art technique such as those proposed in Alvarez-Melis and Saveski (2016) and Huang et al. (2017) can be employed for extracting and modeling active topics. Therefore, we will not be engaged with proposing a new method for the identification of topics and will only focus on determining the interest of users towards the topics once they are identified. Given this focus, we review the work related to the problem of user interest detection from social networks. Interested readers are encouraged to see Aiello et al. (2013), Farzindar and Khreich (2015) and Srijith, Hepple, Bontcheva, and Prentiu-Pietro (2017) for the state of the art on topic and event detection.

Current approaches to user interest identification from social networks can be viewed as either single-source or multi-source (Abel, Herder, Houben, Henze, & Krause, 2013). In single-source approach, only one social network is considered as the source of information. Most of these works use Twitter as their source of information, because the information that the users publish on Twitter are more publicly accessible compared to other social networks. Multi-source approach, on the other hand, is based on the idea that a user has different profiles in different social networks, and to extract her interests more accurately, it would be better to extract and integrate her information from all those profiles (Spasojevic et al., 2014). Independent from how the information is collected and

Download English Version:

<https://daneshyari.com/en/article/6926120>

Download Persian Version:

<https://daneshyari.com/article/6926120>

[Daneshyari.com](https://daneshyari.com)