Semantics-Enabled User Interest Mining

Fattane Zarrinkalam(✉)

Department of Computer Engineering,
Ferdowsi University of Mashhad, Mashhad, Iran
fattane.zarrinkalam@stu-mail.um.ac.ir

Abstract. Microblogging services such as Twitter allow users to express their feelings and views in real-time through microposts. This provides a wealth of information both collectively and individually that can be effectively mined so as to facilitate personalization, recommendation and customized search. A fundamental task with this respect would be to extract users' interests. This has been mainly done using probabilistic models that rely on measures such as frequency of co-occurrence of important phrases, which forgoes the underlying semantics of the phrases in favor of highlighting the role of syntactical repetition of content. Some recent works have considered the role of semantics by using knowledge bases such as DBPedia and Freebase. However, they limit the topics of interest to be a set of individual concepts extracted from the microposts in isolation, i.e. without considering the relationships of the microposts to each other or to other users. This proposal seeks to further build on these works by introducing a definition of topical interest, which enables the identification of more specific and semantically complex topics involving multiple interrelated concepts. Based on this definition, methods will be introduced for the detection of both explicitly observed and implicitly implied user interests, in addition to the identification of user interest shifts based on the temporal clues.

Keywords: #eswcphd2015zarrinkalam · User interest detection · Microblogging service · Semantics-enabled

1 Introduction

With the emergence and the growing popularity of microblogging services like Twitter, many users extensively use microposts to express their feelings and views about different topics. This has made microblogging services a source of implicit and explicit information for user interest identification [1, 2]. This has the potential to contribute to different application areas such as filtering twitter streams [3, 4], news recommendation [5] and user community identification [6], among others.

When processing microposts for the identification and extraction of user interests, traditional keyword-based methods, which are often proposed for processing formal and large documents, are less effective on microposts, due to the short length, noisiness and informality of the content [7, 8]. A potential approach for addressing these issues is to consider the underlying semantics of microposts. To this end, recent works have proposed to utilize external knowledge bases (such as DBpedia) to link the terms in the microposts to the relevant concepts described in those knowledge bases. Since these

F. Gandon et al. (Eds.): ESWC 2015, LNCS 9088, pp. 817–828, 2015.
DOI: 10.1007/978-3-319-18818-8_54

knowledge bases represent the concepts and their relationships, these links provide a way of inferring underlying semantics of the microposts [8–10]. We intend to further build on this approach. The following example provides the basis for this proposal as it distinguishes its contributions from the state of the art.

Motivating Example. Each February, Tim Hortons, a well-known Canadian chain restaurant holds a campaign called *Roll Up the Rim to Win*. A customer can try his luck by buying a paper cup of coffee and unrolling the rim of the cup after finishing his drink, to determine whether he has won a prize, where the greatest one is a Toyota Camry. During the time when the campaign is being held, many users tweet about this event. These tweets contain terms for which a related DBpedia concept, e.g. *Tim Hortons*, *Toyota Camry*, *Roll Up* and *Coffee,* can be identified. These DBpedia concepts can be used to provide semantic information for the corresponding tweets. It is easy to see that a meaningful topic of interest for this example needs to be constructed using a collection of concepts. However, existing works usually represent each interest using one single concept. Therefore, two DBpedia concepts *Tim Hortons* and *Toyota Camry* are considered as two distinct interests. In other words, these approaches cannot infer that a user is interested in a more specific topic, which is actually a combination of multiple related concepts. Further, they often confine users' interests to a set of predefined concepts (e.g. a subset of DBpedia concepts) and therefore interests to recent events such as *Tim Hortons* campaign that are not among that set cannot be discovered on the fly.

This proposal will address these shortcomings by proposing a framework that considers the semantics of microposts with due consideration given to social network structure and the temporal aspects of social content. Our framework is composed of three main components: (1) The extraction of the so-called *topics* in a given time interval, which are built through conjunction of multiple semantic concepts. For instance, during the February, conjunction of DBpedia concepts *Tim Hortons*, *Toyota Camry*, *Roll Up* and *Coffee* might be considered to be a topic of interest. (2) Interest detection for each individual user as it pertains to the extracted *topics*, whether it be explicitly observed or implicitly implied; (3) The temporal modeling of each user's interest shifts with regards to extracted topics.

The rest of the proposal is organized as follows. Section 2 briefly reviews the related work. The problem statement and contributions are presented in Sect. 3, and the proposed approach is introduced in Sect. 4. Section 5 outlines an evaluation plan, and finally, Sect. 6 concludes the proposal.

2 Background Literature

There are three different types of information available on social networks, which have been used in the literature for extracting user interests: (1) User-generated textual contents, such as Twitter posts (*content-based*), (2) Social network structure that shows the relationships between users (*network structure-based*), and (3) Temporal factors that represent the dynamic nature of user interests (*temporal*).

2.1 Content-Based Approaches

There are different approaches for extracting users' interest through the analysis of the user generated textual content. In the *Bag of Words* approach, users' interests are represented as a set of terms extracted from the users' contents [2, 11, 12]. For example, Yang et al. [11] have used a weighted term vector for modeling user interests, and applied cosine similarity for measuring the similarity of users.

Topic Modeling approach provides a probabilistic model for the term frequency occurrences in documents of a given corpus. As a matter of fact this approach forms topics by extracting groups of co-occurring terms and views each document as a mixture of various topics [13]. Latent Dirichlet Allocation (LDA), as a well-known topic modeling method, is frequently used for interest detection [14–16]. For example, Weng et al. [16] have created a single document from the collection of a user's tweets, and then have discovered the topics by running LDA over this document.

Since the *Bag of Words* and *Topic Model* approaches focus on terms without considering their semantic and the relationship between them, they cannot utilize underlying semantics of textual content. Furthermore, these approaches assume that a single document contains rich information, as a result they may not perform so well on short, noisy and informal texts like twitter posts [7–9]. To address these issues, there is another line of work for extracting user interests from microposts through representing user interests as a *Bag of Concepts*. Usually, external knowledge bases such as DBpedia/Wikipedia, Freebase and Yago are used as a source for extracting the candidate concepts. Since these knowledge bases represent the concepts and their relationships, they provide a way of inferring underlying semantics of the content [8–10]. For example, Michelson and Macskassy [8] have proposed Twopics which first extracts a set of Wikipedia entities from a user's tweets and then identifies the high-level interests of the user by traversing and analyzing the Wikipedia categories of the extracted entities. Kapanipathi et al. [3] have modeled users' interests by annotating their tweets with DBPedia concepts, and have used these annotations to filter tweets based on the users' interests. Abel et al. [17] have proposed to enrich twitter messages by linking them to related news articles and then extracting the entities mentioned in the enriched messages as the users' interests. Kapanipathi et al. [9] have introduced two kinds of interests for a user: (1) weighted primitive interests, which is bag of concepts extracted from the entities mentioned in the user's tweets and (2) implicit interests extracted by mapping primitive interests to Wikipedia category hierarchy using a spreading activation algorithm.

2.2 Network Structure-Based Approaches

The social connections of the users are another kind of information that can be used for user interest extraction from social networks [4, 14, 19, 20]. The social connections are usually modeled as a graph in which nodes are users and edges represent their connections. Theory of Homophily [18] is followed by most of the works in this category and it refers to the tendency of users to connect to users with common interests or preferences. For example, Mislove et al. [19] have used this theory to infer missing information and interests of a user based on the information provided by her neighbors. Pennacchiotti et al. [4] have extracted the interests of a user by using tweets of the neighboring users in addition to her own tweets. Wang et al. [14] have extended the Homophily theory by proposing a specific link structure assumption under which local link structures between two nodes are consi-dered to be an indicator of node similarity. For example, if two users share many followers, they are likely to be similar in terms of topical interests.

2.3 Temporal Approaches

Temporal aspects are also considered in some works to infer user interests from social networks [1, 5, 21]. For example Abel et al. [5, 21] have shown that a user's interests change over time and are influenced by public trends. They have modeled user interests in a given timestamp as a set of weighted concepts which are entities or hashtags extracted from the user's tweets in that timestamp. For calculating the weight of each concept, the tweets with shorter temporal distance to the given timestamp are assigned greater weight since they are considered to be more important. The authors have also shown that considering temporal dynamics of the user interests can improve the performance of a personalized news recommender system.

2.4 Discussion

Several interesting works have been performed on extracting users' topical interests from microbloging services. However, the current works struggle with at least one of the following limitations:

- In most studies [1–3, 5, 8–10, 14, 17], each topic of interest is considered to be represented by a single concept. Therefore, it is not possible to infer more specific topics which are only expressible by combining multiple related concepts. Using these approaches, for instance, given a tweet "*Tim Hortons RRRoll Up Replay Game: Tim Hortons RRRoll Up Replay Game Prizes: (1): 2015 Toyota Camry XSE*", may identify *Tim Hortons* and *Toyota Camry* as two distinct topics. The user might not be too interested in *Toyota Camry* as a general topic, but is rather interested in a campaign which includes *Toyota Camry* and *Tim Hortons* together.
- In most studies [1, 3, 5, 8–10, 14, 17], semantic topics of interest are confined to a set of predefined concepts, e.g. only Wikipedia categories, and it is not possible to identify emerging topical interests which are not yet in this predefined initial set. For instance, when an event like *Tim Hortons campaign* appears for the first time, it might rapidly show itself as a topic in the tweets just after a few minutes, but can take much longer to have a Wikipedia page created for it.
- Most of the current works [1, 3, 8–10, 14] do not consider the context of the microposts to extract users' interests. In other words, these works overlook the fact that users usually make an implicit assumption that the readers are aware of the context in which the post is being made. So, understanding the underlying semantics of a post may require consideration of the relationships of posts to each other or to other users. For example, a user might have replied to many tweets related to *Tim Hortons campaign*, without mentioning any of the buzzwords.

- There are some works that consider the temporal aspects for identification of the users' interests [1, 5, 21]. However, they generally do not take into account identification of the user's interest shifts during time, while this is valuable and it can provide valuable insight about the evolution of the users' behavior and distinguishing between his short-term and long-term interests. For instance, knowledge about the interest shifts makes it possible to distinguish between a community of users who show interest in *Tim Hortons* only each February during the campaign and a community of users who follow this topic throughout the year.

3 Problem Statement and Contributions

This proposal seeks to address the limitations discussed in the previous section by proposing a framework that views the content of a social network as a temporal graph. This graph is composed of three heterogeneous vertice types representing (i) individual users, (ii) social contents such as microposts, and (iii) semantic concepts. More specifically, this proposal pursues the following three main contributions:

- We propose to model user interests through a collection of topical interest. We consider each topical interest a conjunction of several coherent semantic concepts. To globally identify so-called *topics* in a given time interval from the social network graph as defined in Sect. 4.1, a concept graph is built in which the vertices represent the semantic concepts extracted from the microposts published in that interval, and the edges indicate semantic relatedness between each two concepts (Sect. 4.2). Each topic is considered to be a cluster in this graph which includes a set of sufficiently related concepts in that time interval. This has the added benefit that each detected interest does not necessarily need to be from amongst a set of predefined concepts, and also, it makes it possible to define semantically complex topics which involve multiple concepts as opposed to single terms or concepts;
- We view a specific user's interests as a set of topics identified from the social network. This set includes explicitly observed interests of the user and also the implicitly implied interests. For a user, the explicit interests are identified from the concepts he has explicitly mentioned in his microposts, with due consideration given to the relationships of microposts to each other or to other users. The implicit interests are the topics that the user is expected to be interested in, and these topics are identified based on the interests of the communities the user is a member of. The proposed framework includes a component for identifying these communities, and based on the identified communities, the implicit interests of the users are determined.
- We further postulate that a user's topical interests can differ and/or evolve based on different time intervals, which refer to as user interest shift. We will propose methods that will be able to accurately model and predict user interest shifts.

4 Proposed Approach

This section describes the underlying representation model of the proposed framework, along with its technical contributions.

4.1 Representation Model

The proposed framework is designed around viewing the data of a microbloging service as a heterogeneous graph with three types of vertices: (1) User vertices representing the individual users. (2) Content vertices representing the contents published by the users. (3) Concept vertices representing the underlying semantics of social contents. Further, the edges of the graph include instances of the different types of relationships between the users, social contents and concepts. It is important to note that in the model not only vertices of the same type can be interconnected, but also different vertices types can be connected to each other.

For instance, in the case of Twitter, as shown in Fig. 1, content vertices include the tweets and the Web pages mentioned in each tweet. Concept vertices can be DBpedia concepts that can be derived directly from the tweets or indirectly from the content of the Web pages mentioned in the tweets. Furthermore, some relationships that can be used include: Follow relation between two users, relation between a user and the tweets she has made or retweeted or marked as 'Favorite', relation between a tweet and the Web pages linked in the tweet, relation between a tweet and the concepts associated with that tweet and others.

The amount of information shown in the network graph of Fig. 1 is readily available in Microblogging services, except for the concept vertices and their associated relationships. These concepts can be extracted using existing systems such as TAGME [22] and DBpedia Spotlight [23] which can be used to annotate a textual content with the resources in Wikipedia/DBpedia. For example, for a given tweet "*Tim Hortons roll up the rim abuses my love for coffee AND gambling*", DBpedia Spotlight identifies three links to DBpedia: *Tim Hortons* is linked to the DBpedia concept represented in "http://dbpedia.org/resource/Tim_Hortons"; *Coffee* is linked to "http://dbpedia.org/page/Coffee" and *Gambling* is linked to "http://dbpedia.org/page/Gambling". The weighted edges between any two concept vertices represent the semantic relatedness of those concepts. This relatedness value generally, not in a specific time interval, can be computed using a Wikipedia-based measure, which for instance computes the relatedness by link structure analysis techniques over wikipedia pages.

To consider the fact that the user interests are not static and they change over time, it is required to represent the network graph as a temporal graph. We will use one of the existing techniques [24, 25] which enable efficient storage and retrieval of temporal graphs and allow retrieving specific snapshots of the network graph. In our proposed approach, time is divided into fixed length intervals and a snapshot of the network graph is retrieved for each time interval $[t_{k-1}, t_k]$. This snapshot includes the users of the social network at time t_k, the contents added to the network during the corresponding time interval, and the concepts associated with these contents.

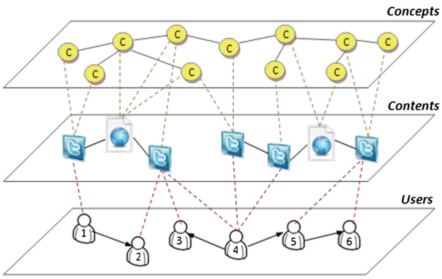

Fig. 1. Representation model

4.2 Concrete Contributions

The proposed framework includes three main contributions which rely on the foundations of the representation model described above. Next, these contributions are described.

Complex Semantic Topic Extraction. The goal of this component is to identify the so-called *topics*, which are modeled by clusters of concepts that are interrelated on the network graph (shown in Fig. 2). To identify these topics for each time interval, it is required to measure the semantic relatedness of the identified concepts in that time interval. Since the semantic relatedness between two concepts changes over time [26], by relying only on the static nature of knowledge bases like DBpedia it is not possible to consider the temporal issues effectively. For instance, computing the relatedness of *Tim Hortons* and *Toyota Camry* based on DBpedia link structure analysis results in the same small value both in February and August. But these concepts may appear so frequently in the users' microposts in February, due to the *Tim Hortones campaign*.

There are some works that seek to address temporal issues by utilizing the dynamics of the social network for computing relatedness of the concepts in a timely manner [26, 27]. However, they compute the relatedness of two concepts in a specific time interval only based on the co-occurrence of those concepts in the microposts published in that time interval. In contrast, we are seeking to provide improvement over these works by considering valuable information reflected in the 3-layer representation model of the network graph. A potential method is discussed as follows.

The relatedness of two concepts *C1* and *C2* at a given time interval can be calculated based on how similar are the content vertices associated with *C1* to the content vertices associated with *C2*. Following the idea of SimRank measure [28], similarity of two content vertices *C1'* and *C2'* can then be computed based on the similarity of the content (user) vertices associated with *C1'* to the content (user) vertices associated with *C2'*. Likewise, similarity of the user vertices can be computed based on the similarity of their associated users and contents.

The overall relatedness of two concepts in a time interval can therefore be computed as a weighted sum of two relatedness values, i.e. the temporal relatedness computed by the method described in the previous paragraph, and the static DBpedia-based relatedness. The weight values are expected to be obtained experimentally.

The computed relatedness values of the concepts are added to the network graph corresponding to a given time interval, in terms of weighted edges between the concepts. Finally, as illustrated in Fig. 2, the topics are determined by applying a graph-based clustering method on the resulting weighted graph.

User Interest Detection. After the topics are identified and modeled from the network graph, individual user's interests are modeled as a function of the identified topics. Our goal is to identify both *explicitly* expressed interests and also *implicitly* inferred interests of each user. To identify explicitly observed interests, we would need to measure the interest of each user against each topic of interest based on the content vertices associated with that user. The basic idea is that the more frequently the concepts of a topic are mentioned in the contents of a user, the more interested the user may be in that topic. We are going to augment this idea with using context information of the user contents. For instance, it is possible that a user has replied to a tweet which is much related to *Tim Hortons campaign*, but the reply itself does not mention any of the concepts associated with this topic. The simple idea mentioned above is unable to see the fact that the reply tweet is also related to that topic, and therefore does not notice the user's interest in the topic.

In order to identify implicitly inferred relations of users to identified topics, it is interesting to extract user-topic communities. As illustrated in Fig. 3, each of these communities include the largest set of mutually similar-enough topics along with the users interested in those topics. To identify these communities, we would need to measure the similarity between each pair of topics. This can be performed by measuring similarity of each topic to a set of predefined high-level topics that can be extracted from existing knowledge bases (e.g. the high-level DBPedia categories). Having the user-topic communities created, the implicitly implied interests of a user can be determined as the topics belonging to the communities in which the user resides in.

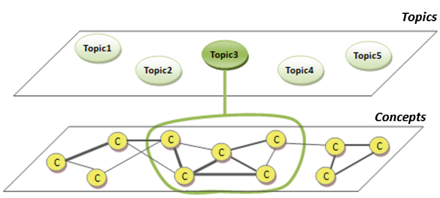

Fig. 2. Topic extraction

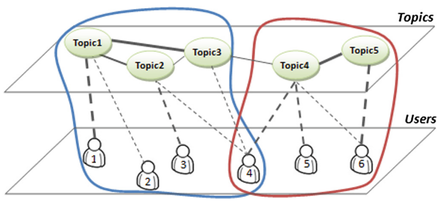

Fig. 3. User-topic communities

Interest Shift Detection. In order to address the interest shift aspect of user interest detection, the identified interests of a user in several consecutive time intervals are monitored. This helps in differentiating between the short-term and the long-term interests of the user. Further, by considering similarity and relatedness of the topic of interests of a user in different intervals, it is possible to model the evolution of the user's behaviour over different topics, i.e. how his interests are attenuated against some topics and focused on some other ones.

One possible way of analyzing interest shift of a user, is port his topics of interest for different time intervals to a set of points in the *2D* space. Knowing that a user is interested in a set of m topics T_k at time interval k, and a set of n topics T_{k+1} at time interval $k+1$, it is possible to build a matrix $m \times n$ representing distances between each pair of topics (T_i, T_j) where $T_i \in T_k$ and $T_j \in T_{k+1}$. This matrix can be computed based on using our topic similarity measure introduced in the previous section. The distance matrix can then be transformed to a set of points in the *2D* space, using Multidimensional Scaling (MDS) methods. Having the topics of interests ported to the *2D* space, it is possible to devise algorithms for identifying user interest shifts by comparing the position of the user's topics of interest in different time intervals.

5 Evaluation Plan

In order to evaluate the proposed methods, we need to first collect a dataset of real-world social network users. Due to widespread use of Twitter and accessibility of its data, a dataset will be created using Twitter data. Since our method is designed to provide improvement over *Bag of Concepts* approach, we are going to compare it with the state of the art works like [5, 8, 9]. Our evaluation plan includes two main approaches: a user study, and an application-based study.

User Study. As it is acknowledged in different works [1, 9], the most reliable and precise way of evaluating the results of interest detection for a user is to ask the same user to verify the results. Then, the user's feedback can be used for measuring quality of the proposed interest detection method. However, User study is costly and its validity is subject to different types of threats which are hard to address in reality. As a result, we will conduct an application-based study to complement the user study.

Application-based. It is possible to evaluate the proposed method by investigating how it affects the performance of an application which works on the basis of the user interests. Similar to [5], we are going to use news recommender application for this purpose. First, a ground truth is built by collecting, for each user, the news articles from BBC or CNN to which the user has explicitly linked in his tweets (or retweets) in a given time interval. Then, a news recommendation algorithm will be used that is able to recommend news articles based on the user's interests identified by our method. By comparing the recommended news with the ones in the ground truth, it is possible to evaluate quality of the recommendations, and therefore determine how successfully the interests have been identified. Traditional Information Retrieval (IR) metrics like P@K and Mean Reciprocal Rank (MRR) can be used for this step.

It must be noted that since our main goal is not to propose a news recommender system, a simple recommender algorithm, like the one used in [5], will be used for this application based evaluation scenario. An additional point is that, instead of using the prepared ground truth, it is also possible to ask the users to judge the recommendations.

The plan described above, evaluates the quality of the proposed interest detection method. However, in order to investigate the importance of the proposed interest shift detection method, a possible approach is to use it for measuring user similarity.

The idea is that considering similarity of the interest shifts of two users is a more accurate way of measuring those users' similarity, compared to simply considering the users' interests at one time interval. If this idea turns out to be valid, then the results of the interest shift detection can contribute to applications that require measuring similarity of the users, for instance content recommenders that employ the collaborative filtering method and hence need to compare users for finding the neighboring users of a specific user.

6 Conclusions

User interest modeling is the basis and core of many services such as recommendation and customization. Due to the popularity of microblogging services like Twitter and the fact that they are considered as a source of implicit and explicit information about the users' interests, recently, user interest detection from microblogging services has been the subject of many researches. We would like to propose a new framework to extract user interests as semantically complex topics composed of multiple interrelated concepts. This framework views data of a microblogging service as a temporal graph with three types of vertices: *(i)* individual users connections; *(ii)* social contents like microposts; and *(iii)* semantic concepts that represent the underlying semantics of the contents. This framework supports the identification of both the observed interests and implicitly implied interests of the user, with due consideration given to the fact that a user's topics of interest may change with time. It is expected that the proposed framework can address shortcomings of the current interest detection approaches that are based on a more limited notion of topical interest. Further, the proposed approach is expected to be able to improve quality of the applications which work on the basis of user interests.

References

1. Budak, C., Kannan, A., Agrawal, R., Pedersen, J.: Inferring user interests from microblogs. Technical Report, MSR-TR-2014-68 (2014)
2. Shin, Y., Ryo, C., Park, J.: Automatic extraction of persistent topics from social text streams. World Wide Web **17**(6), 1395–1420 (2013). Springer
3. Kapanipathi, P., Orlandi, F., Sheth, A., Passant, A.: Personalized filtering of the twitter stream. In: SPIM Workshop at ISWC 2011, pp. 6–13. CEUR-WS (2011)
4. Pennacchiotti, M., Silvestri, F., Vahabi, H., Venturini, R.: Making your interests follow you on twitter. In: 21st ACM International Conference on Information and knowledge management (CIKM 2012), pp. 165–174. ACM (2012)
5. Abel, F., Gao, Q., Houben, G.-J., Tao, K.: Analyzing user modeling on twitter for personalized news recommendations. In: Konstan, J.A., Conejo, R., Marzo, J.L., Oliver, N. (eds.) UMAP 2011. LNCS, vol. 6787, pp. 1–12. Springer, Heidelberg (2011)
6. Palsetia, D., Patwary, M.M., Agrawal, A., Choudhary, A.: Excavating social circles via user interests. J. Soc. Netw. Anal. Min. **4**(1), 1–12 (2014). Springer

7. Sriram, B., Fuhry, D., Demir, E., Ferhatosmanoglu, H., Demirbas, M.: Short text classification in twitter to improve information filtering. In: 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 841–842. ACM, (2010)
8. Michelson, M., Macskassy, S.A.: Discovering users' topics of interest on twitter: a first look. In: 4th Workshop on Analytics for Noisy Unstructured Text Data, pp. 73–80 (2010)
9. Kapanipathi, P., Jain, P., Venkataramani, C., Sheth, A.: User interests identification on twitter using a hierarchical knowledge base. In: Presutti, V., d'Amato, C., Gandon, F., d'Aquin, M., Staab, S., Tordai, A. (eds.) ESWC 2014. LNCS, vol. 8465, pp. 99–113. Springer, Heidelberg (2014)
10. Lu, C., Lam, W., Zhang, Y.: Twitter user modeling and tweets recommendation based on wikipedia concept graph. In: The AAAI 2012 Workshop on Intelligent Techniques for Web Personalization and Recommender Systems (2012)
11. Yang, L., Sun, T., Zhang, M., Davison, B.D.: We know what @you #tag: does the dual role affect hashtag adoption?. In: 21th International Conference on World Wide Web (WWW, 2012), pp. 261–270. ACM (2012)
12. Chen, J., Narin, R., Nelson, L., Bernstein, M., Chi, E.: Short and tweet: experiments on recommending content from information streams. In: 28th international conference on Human factors in Computing Systems (CHI '10), pp. 1185–1194. ACM (2010)
13. Blei, D.: Probabilistic topic models. Commun. ACM **55**(4), 77–84 (2012)
14. Wang, J., Zhao, W.X., He, Y., Li, X.: Infer user interests via link structure regularization. ACM Trans. Intell. Syst. Technol. (TIST) - Special Issue on Linking Social Granularity and Functions **5**(2), 23 (2014). ACM
15. Ramage, D., Dumais, S., Liebling, D.: Characterizing microblogs with topic models. In: 4th International AAAI Conference on Weblogs and Social Media, pp. 130–137 (2010)
16. Weng, J., Lim, E.P., Jiang, J., He, Q.: TwitterRank: finding topic-sensitive influential twitterers. In: 3rd ACM International Conference on Web Search and Data Mining (WSDM 2010), pp. 261–270. ACM (2010)
17. Abel, F., Gao, Q., Houben, G.-J., Tao, K.: Semantic enrichment of twitter posts for user profile construction on the social web. In: Antoniou, G., Grobelnik, M., Simperl, E., Parsia, B., Plexousakis, D., De Leenheer, P., Pan, J. (eds.) ESWC 2011, Part II. LNCS, vol. 6644, pp. 375–389. Springer, Heidelberg (2011)
18. McPherson, M., Smith-Lovin, L., Cook, J.M.: Birds of a feather: homophily in social networks. Annu. Rev. Sociol. **27**(1), 415–444 (2001)
19. Mislove, A., Viswanath, B., Gummadi, K.P., Druschel, P.: You are who you know: inferring user profiles in online social networks. In: 3th ACM International Conference on Web search and Data Mining (WSDM 2010), pp. 251–260. ACM (2010)
20. Abbasi, M.A., Tang, J., Liu, H.: Scalable learning of users' preferences using networked data. In: 25th ACM Conference on Hypertext and Social Media, pp. 4–12. ACM (2014)
21. Abel, F., Gao, Q., Houben, G.J., Tao, K.: Analyzing temporal dynamics in twitter profiles for personalized recommendations in the social web. In: 3rd International Web Science Conference (WebSci 2011), ACM (2011)
22. Ferragina, P., Scaiella, U.: Fast and accurate annotation of short texts with wikipedia pages. J. IEEE Softw. **29**(1), 70–75 (2012). IEEE
23. Mendes, P.N., Jakob, M., Garc´ıa-Silva, A., Bizer, C.: DBpedia spotlight: shedding light on the web of documents. In: I-Semantics 2011, pp. 1–8. ACM (2011)
24. Khurana, U., Deshpande, A.: Efficient snapshot retrieval over historical graph data. In: The IEEE International Conference on Data Engineering, pp. 997–1008. IEEE (2013)

25. Han, W., Miao, Y., Li, K., Wu, M., Yang, F., Zhou, L., Prabhakaran, V., Chen, W., Chen, E.: Chronos: a graph engine for temporal graph analysis. In: EuroSys. ACM (2014)
26. Milikic, N., Jovanovic, J., Stankovic, M.: Discovering the dynamics of terms semantic relatedness through twitter. In: 1st Workshop on #MSM 2011 (2011)
27. Celik, I., Abel, F., Houben, G.-J.: Learning Semantic relationships between entities in twitter. In: Auer, S., Diaz, O., Papadopoulos, G.A. (eds.) ICWE 2011. LNCS, vol. 6757, pp. 167–181. Springer, Heidelberg (2011)
28. Jeh, G., Widom, J.: SimRank: a measure of structural-context similarity. In: 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'02), pp. 538–543. ACM (2002)