JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

SOCIRANK IDENTIFYING AND RANKING REVELENT NEWS TOPICS USING SOCIAL MEDIA FACTORS

Kavitha, Assistant Professor
Computer science & engineering
Vel tech High tech Rangarajan
Dr. Shakuntala engineering college
Chennai, India
kavithaarunachalam@gmail.com

saiteja.T
Computer science & engineering
Vel tech High tech Rangarajan
Dr. Shakuntala engineering college
Chennai, India
saitejatogaru02@gmail.com

Uday kiran.K
Department of Computer Science and Engineering
vel tech high tech Rangarajan
Dr. Shakuntala engineering college
Chennai, India
udayrays87@gmail.com

Vamsi krishna.B
Department of Computer Science and Engineering
vel tech high tech Dr. Rangarajan
Dr. Shakuntala engineering college
Chennai, India
vamsireddyr49@gmail.com

Tharun.C
Department of Computer science and Engineering
vel tech high tech Dr. Rangarajan
Dr. Shakuntala engineering college
Chennai, India
cherukurutharun48@gmail.com

Abstract: *Mass media sources, specifically the news media, have traditionally informed us of daily events. In modern times, social media services such as Twitter provide an enormous amount of user-generated data, which have great potential to contain informative news-related content. For these resources to be useful, we must find a way to filter noise and only capture the content that, based on its similarity to the news media, is considered valuable. However, even after noise is removed, information overload may still exist in the remaining data—hence, it is convenient to prioritize it for consumption. To achieve prioritization, information must be ranked in order of estimated importance considering three factors. First, the temporal prevalence of a particular topic in the news media is a factor of importance, and can be considered the media focus (MF) of a topic. Second, the temporal prevalence of the topic in social media indicates its user attention (UA). Last, the interaction between the social media users who mention this topic indicates the strength of the community discussing it, and can be regarded as the user interaction (UI) toward the topic. We propose an unsupervised framework—SociRank—which identifies news topics prevalent in both social media and the news media, and then ranks them by relevance using their degrees of MF, UA, and UI. Our experiments show that SociRank improves the quality and variety of automatically identified news topics.*

Index Terms—Information filtering, social computing, social network analysis, topic identification, topic ranking.

INTRODUCTION

The mining of valuable information from online sources has become a prominent research area in information technology in recent years. Historically, knowledge that apprises the general public of daily events has been provided by mass media sources, specifically the news media. Many of these news media sources have either abandoned their hardcopy publications and moved to the World Wide Web, or now produce both hard-copy and Internet versions simultaneously. These news media sources are considered reliable because they are published by professional journalists, who are held accountable for their content. On the other hand, the Internet, being a free and open forum for information exchange, has recently seen a fascinating phenomenon known as social media. In social media, regular, non-journalist users are able to publish unverified content and express their interest in certain events. Microblogs have become one of the most popular social media outlets. One microblogging service in particular, Twitter, is used by millions of people around the world, providing enormous amounts of user-generated data. One may assume that this source potentially contains information with equal or greater value than the news media, but one must also assume that because of the unverified nature of the source, much of this content is useless. For social media data to be of any use for topic identification, we must find a way to filter uninformative information and capture only information which, based on its content similarity to the news media, may be considered useful or valuable.The news media presents professionally verified occurrences or events, while social media presents the interests of the audience in these areas, and may thus provide insight into their popularity. Social media services like Twitter can also provide additional or supporting information to a particular news media topic. In summary, truly valuable information may be thought of as the area in which these two media sources topically intersect. Unfortunately, even after the removal of unimportant content, there is still information overload in the remaining news-related data, which must be prioritized for consumption.To assist in the prioritization of news information, news must be ranked in order of estimated importance. The temporal prevalence of a particular topic in the news media indicates that it is widely covered by news media sources, making it an important factor when estimating topical relevance. This factor may be referred to as the MF of the topic. The temporal prevalence of the topic in social media, specifically in Twitter, indicates that users are interested in the topic and can provide a basis for the estimation of its popularity. This factor is regarded as the UA of the topic. Likewise, the number of users discussing a topic and the interaction between them also gives insight into topical importance, referred to as the UI. By combining these three factors, we gain insight into topical importance and are then able to rank the news topics accordingly. [1]. Yang, C., Sun, M., & Liu, Q. (2014). We know what you want to buy: a demographic-based system for product recommendation on microblogs. Proceedings of the 20th ACM international conference on Information and knowledge management. ACM.This paper presents a demographic-based system for product recommendation on microblogs. It explores the use of social media data to understand user preferences and recommend products accordingly. The study provides insights into leveraging social media factors for personalized recommendations. [2]. Mathioudakis, M., & Koudas, N. (2013). TwitterMonitor: trend detection over the Twitter stream. Proceedings of the 2010 ACM SIGMOD International Conference on Management of data. ACM.TwitterMonitor is a system designed for trend detection over the Twitter stream. The paper discusses the challenges and techniques involved in identifying and tracking trending topics on social media. It offers valuable insights into analyzing social media data for real-time trend identification. [3.] Petrović, S., Osborne, M., & Lavrenko, V. (2018)Streaming first story detection with application to Twitter. Human Language Technologies: The 2011 Annual Conference of the North American Chapter of the Association for Computational Linguistics. Association for Computational Linguistics.This paper introduces a method for streaming first story detection, focusing on Twitter data. It highlights the importance of identifying and ranking prevalent news topics in real-time using social media. The study provides a foundation for understanding the dynamics of news propagation on social platforms. [4]. Hong, L., & Davison, B. D. (2020). Empirical study of topic modeling in Twitter. Proceedings of the first workshop on social media analytics. ACM.The authors present an empirical study of topic modeling in Twitter, which aims to extract and analyze prevalent topics in real-time. The paper explores the effectiveness of various topic modeling techniques and their applications in understanding social media conversations. It offers insights into leveraging topic modeling for prevalent news topic identification. [5]. Zhao, W. X., Jiang, J., Weng, J., He, J., Lim, E. P., Yan, H., & Li, X. (2021). Comparing Twitter and traditional media using topic models. European conference on information retrieval. Springer.This research work compares Twitter and traditional media in terms of topic coverage using topic modeling techniques. It analyzes the differences and similarities between the two mediums, shedding light on the role of social media in capturing prevalent news topics. The study contributes to understanding the unique aspects of social media for news identification.[6]. Bhattacharya, P., Das, S., Ghosh, S., & Ganguly, N. (2014). A framework for ranking of microblogs in a network. Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval. ACM.The authors propose a framework for ranking microblogs in a network, emphasizing the importance of considering social relationships in assessing the importance and influence of social media content. The paper provides insights into ranking social media posts and its relevance to prevalent news topic identification. [7.] Lerman, K., & Hogg, T. (2019). Using a model of social dynamics to predict popularity of news. Proceedings of the 19th international conference on World wide web. ACM.This paper introduces a model of social dynamics to predict the popularity of news articles. It highlights the significance of social factors in determining the spread and impact of news topics. The study contributes to understanding the interplay between social media factors and prevalent news topic ranking. [8]. Jiang, L., Yu, M., Zhou, M., Liu, X., & Zhao, T. (2018). Social media-based collective attention modeling for predicting news popularity. IEEE Transactions on Multimedia, 20(4), 986-997.Sociarank, as you mentioned, is a system for identifying and ranking prevalent news topics using social media factors. Here are some keywords and concepts related to Sociarank and its methodology:

Materials and methods

1. Sociarank: The name of the system that identifies and ranks prevalent news topics using social media factors.2. News topics: Subjects or events that are currently being discussed in the news or on social media platforms.3. Prevalence: The degree to which a news topic is widespread or popular on social media.4. Social media factors: Various metrics or indicators derived from social media platforms that contribute to the ranking of news topics. These factors can include the number of mentions, shares, likes, comments, and overall engagement related to a particular news topic.5. Trending: News topics that are currently popular or gaining significant attention on social media platforms.6. Virality: The extent to which a news topic spreads rapidly and widely across social media networks.7. Sentiment analysis: The process of determining the general sentiment or emotional tone associated with a particular news topic on social media. It can help understand whether the sentiment is positive, negative, or neutral.8. Influence: The impact or authority of social media accounts or users discussing a specific news topic. Influential accounts or users can significantly shape the popularity and perception of news topics. Analysis, Algorithm: The computational process or set of rules used by Sociarank to analyze and rank news topics based on social media factors.10. Real-time monitoring: The continuous tracking and analysis of social media platforms to capture the latest trends and news topics as they emerge.11. Data mining: The process of extracting relevant information or patterns from large datasets, in this case, social media data, to identify prevalent news topics.12. Machine learning: The use of algorithms and statistical models that enable a system like Sociarank to learn from and make predictions or classifications based on social media data.By considering these keywords and concepts, Sociarank aims to provide insights into the most prevalent and influential news topics circulating on social media platforms at any given time.

Proposed Work

The main research areas applied in this paper include: topic identification, topic ranking social, network analysis, keyword extraction, co-occurrence similarity measures, and graph clustering. Extensive work has been conducted in most of these areas.

A. **Topic Identification**

Much research has been carried out in the field of topic identification—referred to more formally as topic modeling. Two traditional methods for detecting topics are LDA [1] and PLSA [2], [3]. LDA is a generative probabilistic model that can be applied to different tasks, including topic identification. PLSA, similarly, is a statistical technique, which can also be applied to topic modeling. In these approaches, however, temporal information is lost, which is paramount in identifying prevalent topics and is an important characteristic of social media data. Furthermore, LDA and PLSA only discover topics from text corpora; they do not rank based on popularity or prevalence. Wartena and Brussee [4] implemented a method to detect topics by clustering keywords. Their method entails the clustering of keywords—based on different similarity measures—using the induced k-bisecting clustering algorithm [5]. Although they do not employ the use of graphs, they do observe that a distance measure based on the Jensen–Shannon divergence (or information radius [6]) of probability distributions performs well. More recently, research has been conducted in identifying topics and events from social media data, taking into account temporal information. Cataldi et al. [7] proposed a topic detection technique that retrieves real-time emerging topics from Twitter. Their method uses the set of terms from tweets and model their life cycle according to a novel aging theory.Additionally, they take into account social relationships—morespecifically, the authority of the users in the network—to determine the importance of the topics. Zhao et al. [8] carried out similar work by developing a Twitter-LDA model designed to identify topics in tweets. Their work, however, only considers the personal interests of users, and not prevalent topics at a global scale.Another trending area of related research is the detection of "bursty" topics (i.e., topics or events that occur in short, sudden episodes). Diao et al. [9] proposed a method that uses a state machine to detect bursty topics in microblogs. Their method also determines whether user posts are personal or refer to a particular trending topic. Yin et al. [10] also developed a model that detects topics from social media data, distinguishing between temporal and stable topics. These methods, however, only use data from microblogs and do not attempt to integrate them with real news. Additionally, the detected topics are not ranked by popularity or prevalence.

B. **Topic Ranking**

Another major concept that is incorporated into this paper is topic ranking. There are several means by which this task can be accomplished, traditionally being done by estimating how frequently and recently a topic has been reported by mass media. Wang et al. [11] proposed a method that takes into account the users' interest in a topic by estimating the amount of times they read stories related to that particular topic. They refer to this factor as the UA. They also used an aging theory developed by Chen et al. [12] to create, grow, and destroy a topic. The life cycles of the topics are tracked by using an energy function. The energy of a topic increases when it becomes popular and it diminishes over time unless it remains popular. We employ variants of the concepts of MF and UA

to meet our needs, as these concepts are both logical and effective. Other works have made use of Twitter to discover news-related content that might be considered important. Sankara Narayanan et al. [13] developed a system called Twitter Stand, which identifies tweets that correspond to breaking news. They accomplish this by utilizing a clustering approach for tweet mining. Phelan et al. [14] developed a recommendation system that generates a ranked list of news stories. News are ranked based on the co-occurrence of popular terms within the users' RSS and Twitter feeds. of these systems aim to identify emerging topics, but give no insight into their popularity over time. Moreover, the work by Phelan et al. [14] only produces a personalized ranking (i.e., news articles tailored specifically to the content of a single user), rather than providing an overall ranking based on a sample of all users. Nevertheless, these works provide us with a basis for extending the premise of UA.Research has also been carried out in topic discovery and ranking from other domains. Shobhakar et al. [15] developed an algorithm that detects and ranks topics in a corpus of research papers. They used closed frequent keyword-sets to form topics and a modification of the PageRank [16] algorithm to rank them. Their work, however, does not integrate or collaborate with other data sources, as accomplished by SociRank.

C. **Social Network Analysis**

In the case of UA, Wang et al. [11] estimated this factor by using anonymous website visitor data. Their method counts the amount of times a site was visited during a particular period of time, which represents the UA of the topic to which the site is related. Our belief, on the other hand, is that, although website usage statistics provide initial proof of attention, additional data are needed to corroborate it. We employ the use of social media, specifically Twitter, as a means to estimate UA. When a user tweets about a particular topic, it signifies that the user is interested in the topic and it has captured her attention more so than visiting a website related to it. In summary, visiting a website might be the initial stimulus, but taking the additional step of discussing a topic via social media signifies genuine attention.1 Additionally, we believe that the relationship between social

media users who discuss the same topics also plays a key role in topic relevance. Kwan et al. [17] proposed a measure referred to as reciprocity, which attempts to detect the interaction between social media users and perceive their engagement in relation to a particular topic. Higher reciprocity means greater interaction between users, and thus topics with higher reciprocity should be considered more important because of their underlying community structure. We can inherently identify the power and influence of a well-structured community asopposed to a decentralized and unstructured one. Our method applies this logic to support the idea that higher reciprocity signifies greater importance.

D. **Keyword Extraction**

Concerning the field of keyword or informative term extraction, many unsupervised and supervised methods have been proposed. Unsupervised methods for keyword extraction rely solely on implicit information found in individual texts or in a text corpus. Supervised methods, on the other hand, make use of training datasets that have already been classified. Among the unsupervised methods, there are those that employ statistical measures of term informativeness orrelevance, such as term specificity [18], TFIDF [19], word frequency [20], n-grams [21], and word co-occurrence [22]. Other unsupervised approaches are graph-based, where a text is converted into a graph whose nodes represent text units (e.g., words, phrases and sentences) and whose edges represent the relationships between these units. The graph is then recursively iterated and relevance scores are assigned to each node using different approaches. A popular example of a graphbased keyword extraction method is TextRank, proposed byMihalcea and Tarau [23]; it utilizes the premise of the popular PageRank [16] algorithm.There has also been much work on keyword extraction using supervised and hybrid approaches. Two traditional supervised frameworks are KEA [24] and GenEx [25], which use machine learning algorithms for the effective extraction of keywords. Other innovative approaches for keyword extraction have been proposed in recent years, including the application of neural networks [26]–[28] and conditional random fields [29]. Hybrid methods (i.e., methods that make use of unsupervised and supervised components) have been proposed as well, such as HybridRank [30], which makes use of collaboration between the two approaches. Due to its simple implementation, we use TextRank [23] to extract keywords from the news media sources. Furthermore, TextRank does not require training or any document corpus for its operation.

E. Co-Occurrence Similarity Matsuo and Ishizuka [22] suggested that the co-occurrence relationship of frequent word pairs from a single document may provide statistical information to aid in the identification of the document's keywords. They proposed that if the probability distribution of co-occurrence between a term x and all other terms in a document is biased to a particular subset of frequent terms, then term x is likely to be a keyword. Even though our intention is not to employ co-occurrence for keyword extraction, this hypothesis emphasizes the importance of co-occurrence relationships. Chen et al. [31] proposed a novel co-occurrence similarity measure in which they measure the association of terms using snippets returned by Web searches. They refer to this measure as co-occurrence double checking (CODC). Bollegala et al. [32] proposed a method that uses page counts and text snippets from Web searches to measure the similarity between words or entities. They compared their method with CODC [31], as well as with variants of several other co-occurrence similarity measures, such as the overlap (Simpson) [33], Dice [34], point-wise mutual information (PMI) [35], Jaccard [36], and cosine similarities.Since establishing the importance of the word-pair cooccurrence distribution in the actual corpus of tweets is of more interest to us, we did not employ Bollegala's or Chen's semantic similarity methods. In this paper, we tested other similarity measures, and found that the Dice similarity measure provided the best results.

F. Graph Clustering

The main purpose of graph clustering in this paper is to identify and separate TCs, as done in Wartena and Brussee's work [4]. Iwasaka and Tanaka-Ishii [37] also proposed a method that clusters a co-occurrence graph based on a graph measure known as transitivity. The basic idea of transitivity is that in a relationship between three elements, if the relationship holds between the first and second elements and between the second and third elements, it also holds between the first and third elements. They suggested that each output cluster is expected to have no ambiguity, and that this is only achieved when the edges of a graph (representing co-occurrence relations) are transitive. Matsuo et al. [38] employed a different approach to achieve the clustering of co-occurrence graphs. They used Newman clustering [39] to efficiently identify word clusters. The core idea behind Newman clustering is the concept of edge betweenness. The betweenness measure of an edge is the number of shortest paths between pairs of nodes that run along it. If a network contains clusters that are loosely connected by a few intercluster edges, then all shortest paths between different clusters must go along one of these edges. Consequently, the edges connecting different clusters will have high edge betweenness, and removing them iteratively will yield well-defined clusters.

SOCIRANK FRAMEWORK

The goal of our method—SociRank—is to identify, consolidate and rank the most prevalent topics discussed in both news media and social media during a specific period of time. The system framework can be visualized in Fig. 1. To achieve its goal, the system must undergo four main stages.

1) Preprocessing: Key terms are extracted and filtered from news and social data corresponding to a particular period of time.2) Key Term Graph Construction: A graph is constructed from the previously extracted key term set, whose vertices represent the key terms and edges represent the co-occurrence similarity between them. The graph, after processing and pruning, contains slightly joint clusters of topics popular in both news media and social media.3) Graph Clustering: The graph is clustered in order to obtain well-defined and disjoint TCs.4) Content Selection and Ranking: The TCs from the graph are selected and ranked using the three relevance factors (MF, UA, and UI). Initially, news and tweets data are crawled from the Internet and stored in a database. News articles are obtained from specific news websites via their RSS feeds and tweets are crawled from the Twitter public timeline [41]. A user then requests an output of the top k ranked news topics for a specified period of time between date d1 (start) and date d2 (end).A. Preprocessing In the preprocessing stage, the system first queries all news articles and tweets from the database that fall within date d1 and date d2. Additionally, two sets of terms are created: one for the news articles and one for the tweets, as explained below.

1) News Term Extraction: The set of terms from the news data source consists of keywords extracted from all the queried articles. Due to its simple implementation and effectiveness, we implement a variant of the popular Text Rank algorithm [23] to extract the top k keywords from each news article. 2 The selected keywords are then lemmatized using the WordNet lemmatizer in order to consider different inflected forms of a word as a single item. After lemmatization, all unique terms are added to set N. It is worth pointing out that, since N is a set, it does not contain duplicate terms.2) Tweets Term Extraction: For the tweets data source, the set of terms are not the tweets' keywords, but all unique and relevant terms. First, the language of each queried tweet is identified, disregarding any tweet that is not in English. From the remaining tweets, all terms that appear in a stop word.

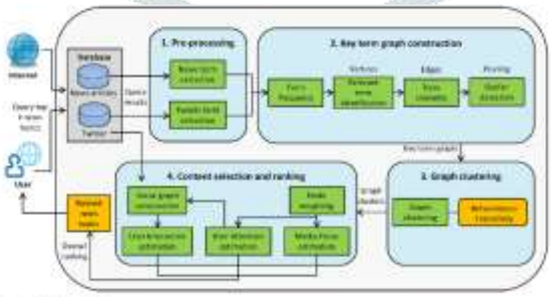
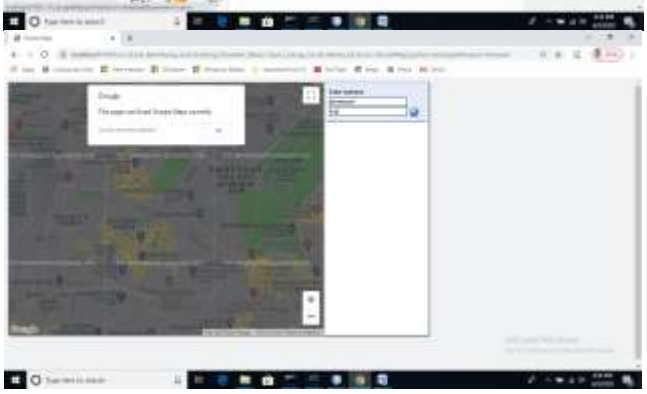
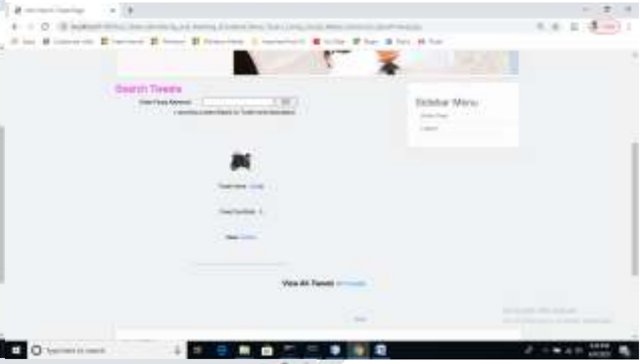


Fig. 1. SociRank framework.

list or that are less than three characters in length are eliminated. The part of speech (POS) of each term in the tweets is then identified using a POS tagger [42]. This POS tagger is especially useful because it can identify Twitter-specific POSs, such as hashtags, mentions, and emoticon symbols. Hashtags are of great interest to us because of their potential to hold the topical focus of a tweet. However, hashtags usually contain several words joined together, which must be segmented in order to be useful. To solve this problem, we make use of the Viterbi segmentation algorithm [43]. The segmented terms are then tagged as "hashtag." To eliminate terms that are not relevant, only terms tagged as hashtag, noun, adjective or verb are selected. The terms are then lemmatized and added to set T, which represents all unique terms that appear in tweets from dates d1 to d2.B. Key Term Graph ConstructionIn this component, a graph G is constructed, whose clustered nodes represent the most prevalent news topics in both news and social media. The vertices in G are unique terms selected from N and T, and the edges are represented by a relationship between these terms. In the following sections, we define a method for selecting the terms and establish a relationship between them. After the terms and relationships are identified, the graph is pruned by filtering out unimportant vertices and edges.





CONCLUSION

In this paper, we proposed an unsupervised method— SociRank—which identifies news topics prevalent in both social media and the news media, and then ranks them by taking into account their MF, UA, and UI as relevance factors. The temporal prevalence of a particular topic in the news media is considered the MF of a topic, which gives us insight into its mass media popularity. The temporal prevalence of the topic in social media, specifically Twitter, indicates user interest, and is considered its UA. Finally, the interaction between the social media users who mention the topic indicates the strength of the community discussing it, and is considered the UI. To the best of our knowledge, no other work has attempted to employ the use of either the interests of social media users or their social relationships to aid in the ranking of topics. Consolidated, filtered, and ranked news topics from both professional news providers and individuals have several benefits. One of its main uses is increasing the quality and variety of news recommender systems, as well as discovering hidden, popular topics. Our system can aid news providers by providing feedback of topics that have been discontinued by the mass media, but are still being discussed by the general population. SocialRank can also be extended and adapted to other topics besides news, such as science, technology, sports, and other trends. We have performed extensive experiments to test the performance of SocialRank, including controlled experiments for its different components. SocialRank has been compared to media focus-only ranking by utilizing results obtained from a manual voting method as the ground truth. In the voting method, 20 individuals were asked to rank topics from specified time periods based on their perceived importance. The evaluation provides evidence that our method is capable of effectively selecting prevalent news topics and ranking them based on the three previously mentioned measures of importance. Our results present a clear distinction between ranking topics by MF only and ranking them by including UA and UI. This distinction provides a basis for the importance of this paper, and clearly demonstrates the shortcomings of relying solely on the mass media for topic ranking. As future work, we intend to perform experiments and expand SocialRank on different areas and datasets. Furthermore, we plan to include other forms of UA, such as search engine click-through rates, which can also be integrated into our method to provide even more insight into the true interest of users. Additional experiments will also be performed in different stages of the methodology. For example, a fuzzy clustering approach could be employed in order to obtain overlapping TCs (Section III-C). Lastly, we intend to develop a personalized version of SocialRank, where topics are presented differently to each individual user.

REFERENCES

[1] D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent Dirichlet allocation,” J. Mach. Learn. Res., vol. 3, pp. 993–1022, Jan. 2003.[2] T. Hofmann, “Probabilistic latent semantic analysis,” in Proc. 15th Conf. Uncertainty Artif. Intell., 1999, pp. 289–296.[3] T. Hofmann, “Probabilistic latent semantic indexing,” in Proc. 22nd Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, Berkeley, CA, USA, 1999, pp. 50–57.[4] C. Wartena and R. Brussee, “Topic detection by clustering keywords,” in Proc. 19th Int. Workshop Database Expert Syst. Appl. (DEXA), Turin, Italy, 2008, pp. 54–58.[5] F. Archetti, P. Campanelli, E. Fersini, and E. Messina, “A hierarchical document clustering environment based on the induced bisecting k-means,” in Proc. 7th Int. Conf. Flexible Query Answering Syst., Milan, Italy, 2006, pp. 257–269. [Online]. Available: http://dx.doi.org/10.1007/11766254_22.

[6] C. D. Manning and H. Schütze, Foundations of Statistical Natural Language Processing. Cambridge, MA, USA: MIT Press, 1999.[7] M. Cataldi, L. Di Caro, and C. Schifanella, “Emerging topic detection on Twitter based on temporal and social terms evaluation,” in Proc. 10th Int. Workshop Multimedia Data Min. (MDMKDD), Washington, DC, USA, 2010, Art. no. 4. [Online]. Available: http://doi.acm.org/10.1145/1814245.1814249.[8] W. X. Zhao et al., “Comparing Twitter and traditional media using topic models,” in Advances in Information Retrieval. Heidelberg, Germany: Springer Berlin Heidelberg, 2011, pp. 338–349.

[9] Q. Diao, J. Jiang, F. Zhu, and E.-P. Lim, “Finding bursty topics from microblogs,” in Proc. 50th Annu. Meeting Assoc. Comput. Linguist. Long Papers, vol. 1. 2012, pp. 536–544.

[10] H. Yin, B. Cui, H. Lu, Y. Huang, and J. Yao, “A unified model for stable and temporal topic detection from social media data,” in Proc. IEEE 29th Int. Conf. Data Eng. (ICDE), Brisbane, QLD, Australia, 2013, pp. 661–672.[11] C. Wang, M. Zhang, L. Ru, and S. Ma, “Automatic online news topic ranking using media focus and user attention based on aging theory,” in Proc. 17th Conf. Inf. Knowl. Manag., Napa County, CA, USA, 2008,pp. 1033–1042.[12] C. C. Chen, Y.-T. Chen, Y. Sun, and M. C. Chen, “Life cycle modeling of news events using aging theory,” in Machine Learning: ECML 2003. Heidelberg, Germany: Springer Berlin Heidelberg, 2003, pp. 47–59.[13] J. Sankaranarayanan, H. Samet, B. E. Teitler, M. D. Lieberman, and J. Sperling, “TwitterStand: News in tweets,” in Proc. 17th ACM SIGSPATIAL Int. Conf. Adv. Geograph. Inf. Syst., Seattle, WA, USA, 2009, pp. 42–51.[14] O. Phelan, K. McCarthy, and B. Smyth, “Using Twitter to recommend real-time topical news,” in Proc. 3rd Conf. Recommender Syst., New York, NY, USA, 2009, pp. 385–388.[15] K. Shubhankar, A. P. Singh, and V. Pudi, “An efficient algorithm for topic ranking and modeling topic evolution,” in Database Expert Syst. Appl., Toulouse, France, 2011, pp. 320–330.[16] S. Brin and L. Page, “Reprint of: The anatomy of a large-scale hypertextual web search engine,” Comput. Netw., vol. 56, no. 18, pp. 3825–3833, 2012.[17] E. Kwan, P.-L. Hsu, J.-H. Liang, and Y.-S. Chen, “Event identification for social streams using keyword-based evolving graph sequences,” in Proc. IEEE/ACM Int. Conf. Adv. Soc. Netw. Anal. Min., Niagara Falls, ON, Canada, 2013, pp. 450–457.[18] K. Kireyev, “Semantic-based estimation of term informativeness,” in Proc. Human Language Technol. Annu. Conf. North Amer. Chapter Assoc. Comput. Linguist., 2009, pp. 530–538.