



HAL
open science

Exploring Big Data Environment for Conversation Data Analysis and Mining on Microblogs

R Belkaroui, D D Jemal, R Faiz

► **To cite this version:**

R Belkaroui, D D Jemal, R Faiz. Exploring Big Data Environment for Conversation Data Analysis and Mining on Microblogs. The 1st International Conference on Advanced Intelligent System and Informatics (AISI2015), Nov 2015, Beni Suef, Egypt. pp.335 - 346, 10.1007/978-3-319-26690-9_30 . hal-01418816

HAL Id: hal-01418816

<https://hal.science/hal-01418816>

Submitted on 21 Dec 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Exploring Big Data Environment for Conversation Data Analysis and Mining on Microblogs

Rami Belkaroui¹, Dhouha Jemal¹, and Rim Faiz²

¹ LARODEC, ISG Tunis, University of Tunis
Bardo, Tunisia

`rami.belkaroui@gmail.com`

`Dh.jemal@gmail.com`

² LARODEC, IHEC Carthage, University of Carthage
Carthage Presidency, Tunisia

`rim.faiz@ihec.rnu.tn`

Abstract. Today, social media services and multiplatform applications such as microblogs, forums and social networks gives people the ability to communicate, interact and generate content which establish social and collaborative backgrounds. These services now embodies the leading and biggest repository containing millions of Big social Data that can be useful for many applications such as measure public sentiment, trends monitoring, reputation management and marketing campaigns. But social media data are essentially unstructured that's what makes it so interesting and so hard to analyze. Making sense of it and understanding what it means will require all new technologies and techniques, including the emerging field of big data. In addition, social media is a key model of the velocity and variety which are main characteristics of Big Data.

In this paper, we propose a new approach to retrieve conversation on microblogging sites that combine Big Data environment and social media analytics solutions. The goal of our approach is to present a more informatives result and solve the information overload problem within Big Data environment. The proposed approach has been implemented and evaluated by comparing it with Google and Twitter Search engines and we obtained very promising results.

Keywords: Social Media, Big Data analysis, data mining, conversation retrieval, social networks analytics.

1 Introduction

Social media is considered as a part of the Web 2.0 movement, which gave birth to a huge volume of data produced by users called User Generated Content (UGC). This massive explosion of volume and data types' diversity implies the imposition of the Big Date in our technological landscape. Furthermore, many of the most important sources of Big Data are relatively new. The huge amount of information from social networks, are only as old as the networks themselves; for example Facebook was launched in 2004, Twitter in 2006. In addition, the term Big data [24] has been mainly used in two contexts, firstly as a technological challenge when dealing with data-intensive domains such as high energy physics, astronomy or internet search, and secondly as a sociological problem when data about us is collected and mined by companies such as Facebook, Google, mobile phone companies, retail chains and governments.

In the current era, social media services attract more and more users and tend to become a solid media for simplified collaborative communication due to the ease and the speed of information sharing especially in real time. Thus, social media services have completely changed the manner in which people communicate and share information. They [13] give people the ability not only to communicate, interact and collaborate with each other, reply to messages from others and create conversations. In addition, social media such as Twitter and Facebook represents an enormous public archive of human thought that captures the ideas, opinions and debates taking place around the world on almost any topic at any moment in time. The analysis of those communications can be

useful in various ways such as measuring public sentiment, trend monitoring, reputation management, marketing campaigns, customer behavior and news analysis. Additionally, the analysis of Big social Data has become substantially to help businesses gain a better understanding of their consumers by incorporating social feedback, which should improve decision-making process. Although, according to Salesforce research, 89% of business leaders believe that Big Data will revolutionize business operations in the same way the Internet did, and 83% have pursued Big Data projects in order to seize a competitive edge.³ As Big Data is becoming a hot topic in many areas where datasets are so large that they can no longer be handled effectively or even completely, it presents an opportunity to analyze expressions and behavioral traces available on social platforms. The question that arises is how to benefit the e-identity and user's opinions shared on social medias.

In this paper, we propose a new approach to retrieve conversation on microblogging sites that combine Big Data environment and social media analytics solutions. The proposed approach can be used to extract conversation from social media in order to provide relevant results corresponding to user's information needs based on their interactions analysis. In particular, the novelty of our approach is the ability to provide an informative relevant data to satisfy any information need in a large social Big Data. The remainder of this paper is organized as follows. In section 2, we give an overview of related work addressing the domains used in our work: Big Data analysis and social media analytics solutions. In section 3, we present our proposed approach for conversation retrieval, then in section 4 we discuss the obtained results. Finally, section 5 concludes this paper and outlines our future work.

2 Related Work

Basically, our work lies at the intersection of two main domains: Big Data analysis and social media retrieval. This section is devoted to presenting these latter.

2.1 Big Data Analysis

Currently, a strong interest towards the term Big Data is arising in the literature. Several research works have focused on this actual research trend in this field. In this context, Sagioglu et al [22] presented an overview of big data's content, scope, samples, methods, advantages and challenges, detail Big Data its main components and discuss privacy concern. In [24], the authors have performed analysis on Flickr, Facebook and Google+ social media sites. Based on this analysis, they have discussed the privacy implications and also geo-tagged social media; an emerging trend in social media sites. The proposed concept in this paper helps users to get informed about the data relevant to them in such large social Big Data. In [9], open problems and actual research trends are highlighted with the aim of providing an overview of state of the art research issues and achievements in the field of analytics over Big Data, and extend the discussion to analytics over big multidimensional data. This work presented several novel research directions arising in this field, which plays a leading role in next-generation Data analysis.

According to [21], the expressions and behavioral traces from large numbers of individuals or groups on social platforms constitute Big Data. The authors proposed to analyze this data as an opportunity to provide valuable insights into the arrays of meaning and practice that emerge and manifest social platforms users online.

In [4], authors presented the problem of misleading claims to objectivity and accuracy for Big Data. They supposed that in the case of social media data, there is a 'data cleaning' process: making decisions about what attributes and variables will be counted, and which will be ignored. They described this process as inherently subjective. In addition to this question, there is the issue of data errors. Authors considered the large data sets from Internet sources as unreliable, prone to outages and losses, and these errors and gaps are magnified when multiple data sets are

³ <http://www.forbes.com/sites/louiscolumnbus/2015/05/25/roundup-of-analytics-big-data-business-intelligence-forecasts-and-market-estimates-2015/>

used together. For this purpose, [3] explains that as a large mass of raw information, Big Data is not self-explanatory. The authors wonder if the data can represent an 'objective truth' or is any interpretation necessarily biased by some subjective filter or the way that data is 'cleaned?'

2.2 Social Media Retrieval and Analysis

Extracting information from Social media Sites is one of the critical problems which have already been addressed over the past years due to its theoretical and practical significance. Social media have become a digital place where users can discuss public issues, share critical information during natural disasters and it will be possible to comment TV shows they are following. Users examine social media sites for answers, timely information (e.g., news, events), people information and topical information [10,26]. However, search functionality provided by those sites is limited to keyword based retrieval to return the most recent posts. However, users are able neither to explore the results nor retrieve more relevant information based on the content [1], and may get lost or become frustrated by the information overload [2].

Recently, various researches have focused on these phenomena with a closer perspective [5,18,20]. In [19] the authors proposed a user-based tree model for retrieving conversations from microblogs. They only considered tweets that directly respond to other tweets by using the "@username" as a marker of addressivity. The advantage of this method is having a coherent conversation based on the direct links between users. The downside of this method is the fact of neglecting the tweet that does not contain the "@sign". Nevertheless, we can have tweets that are related to the conversation even though they are not directly linked with other tweets. In [8], the authors proposed a method to build conversation graphs. The authors focused on the particular case of a conversation formed by users replying to tweets. In this case, a tweet can only reply to one other tweet and a retweet are ignored, but users can get involved in conversations with other by commenting, liking and sharing other user's posts. Other related works concentrated on different aspects of microblogging conversations are [12,25], the deal respectively with the tagging of conversations and the identification of topics. Kumar et al [15], in their work has proposed a simple mathematical model that produces basic conversation structures taking into account the identities of each member of the conversation.

Besides, some recent studies aimed to identify how information flow depends on the structural properties of user interactions [6,16]. Moreover, modeling the interaction among users plays a crucial role in order to understand how the information is disseminated in the network [27,11] and how to maximize its spread [14,7]. In [16], the authors proposed the first large scale analysis of both topological and temporal structures of retweets in order to catch users' popularity of, trending topics and temporal patterns describing the propagation of information in the network. The temporal dimension of the information spread plays an important role in characterizing users' retweets and identifying influential as well [17,28]. Eventually, we have to mention that in dynamic interactions, social cascades have also been used to reveal that, different from static properties, the user's geographic locality is a key factor to characterize the way how the information circulates in the network [23].

In summary, retrieving information from social media sites represented always a challenge given its volume, inconsistent writing and noise. Most existing systems focus on term-based approach, but retrieving the relevant information is often neglected, leading to less satisfactory results while searching information.

3 Conversations Extraction Approach on Big Twitter Data

In recent years, conversations generated by Big Data environments (Twitter, Facebook, blogs, etc.) have flooded the web. Big Social Data have transformed the scale of exploratory analysis on the web and offered new means of performing tasks that were not feasible before. This highlighted how Big Data research has become central to this growth of knowledge. Conversations have become an essential feature of contents generated by most social media platforms. Particularly, on Twitter, many users post tweets freely in order to express what they are thinking about any event, topic,

followed by some comments, retweet or favorite. People uses conversations to express their interests, feelings, and experiences about virtually any topic. Thus, the users' interactions essentially reflect the importance of different topics and can be used to improve the conversation retrieval task quality.

In this paper, we explore the Big Data environment to propose an information retrieval approach for microblogging sites. Particularly, we focused on Twitter based conversation retrieval task (figure 1). Our approach combines direct and indirect conversation aspects in order to extract extensive posts beyond conventional conversation. In addition, we defined a conversation as a set of short text messages posted by a user at specific timestamp on the same topic. These messages can be directly replied to other users by using "@username" or indirectly by liking, retweeting, commenting and other possible interactions.

For this purpose, we opted for a Big Data solution. Due to the huge amount of data, structure diversity and specially the task complexity, a Big Data solution can provide best performance and efficiency. Thus, MapReduce is presented as one of the most efficient Big Data solutions. The MapReduce programming model has been successfully used for many different purposes. Hence, the remedy of the user defined function, which is one of the main strength feature of the MapReduce model. While conducting our exploratory detection and analysis of the tweeter's conversation, we came across large conversations' features. Our approach is bases on the direct and indirect conversations and consist in 2 steps:

- Step 1: Constructing the direct reply tree using all tweets in reply directly to other tweets.
- Step 2: Detecting the relevant tweets related indirectly to a same reply tree which might be retweets, comments or other possible interactions in order to extract extensive posts beyond conventional conversation.

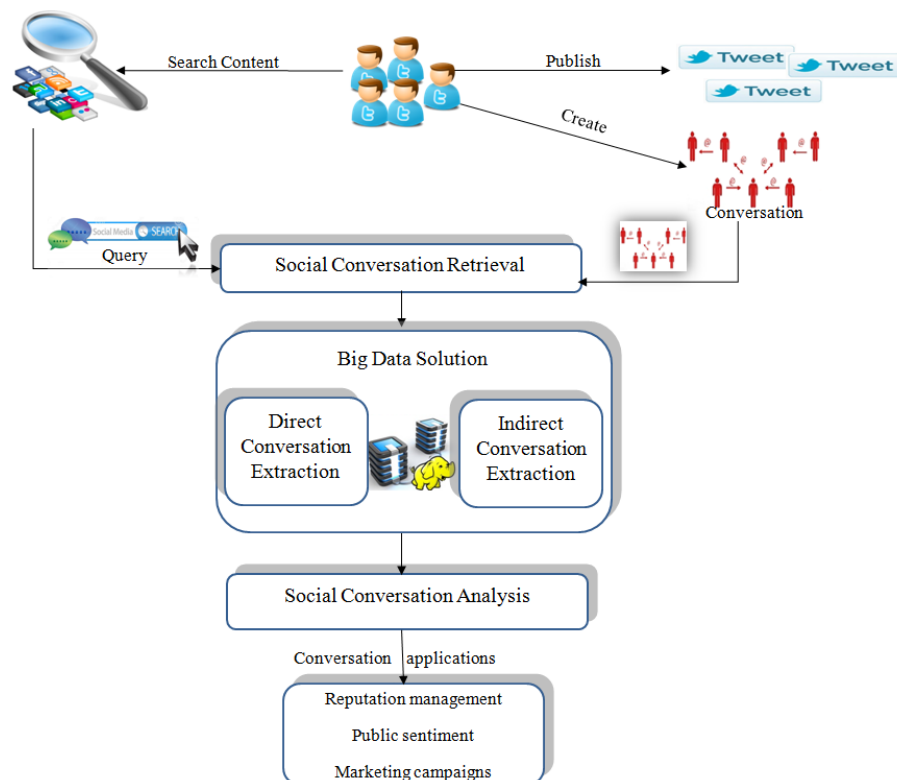


Fig. 1. Exploring Big Data environment for Twitter conversations extraction approach

3.1 Twitter Conversations Construction Using MapReduce

In this phase, we aim to collect all tweets in reply directly to other tweets. Obviously, a reply to a user will always begin with "@username". Our goal in this step is to create reply tree. The reply tree construction process consists of two algorithms: recursive root finder and iterative search algorithm.

For this purpose, we implement the two MapReduce phases: The map phase is implemented with a recursive algorithm to find T0 the root (first tweet published) of the conversation and H0 the hashtag used in the selected conversation. Thus, while the type of a given tweet is a reply or a retweet, the recursive algorithm boots until identifying the root. The iterative algorithm runs in parallel to seek the remainder of the conversation and extracting the used hashtag. The established conversations and the used hashtags, will be the input of the reduce phase of the MapReduce process. The goal of this phase is to classify conversations based on identical hashtags, using a grouping conversations algorithm.

3.2 Indirect Reply Structure Using Conversational Features

To the best of our knowledge, there has not been previous work on the structure of reply-based on indirectly conversation. In this step, we also used the MapReduce model to enrich conversations established during the first step. The goal of this step is to extract tweets that may be relevant to the conversation without the use of the @symbol. The conversations classification based on hashtags will facilitate the selection of the candidate tweets for each conversations cluster. Therefore, we define new features that may help to detect tweets related indirectly to a same conversation. The features are used during the MapReduce process. The Map phase is designed to calculate a score for each tweet relative to each conversation. The conversations are the output of the first step of our proposed approach. The features we used are:

- *URLs-based selection:*

Twitter allows users to include URLs as supplement information to their tweets. URLs that are contained in tweets can be considered as indicators for news related tweets. By sharing an URLs, an author would enrichment the information published in his tweet. In particular, if a tweet contains a URL that points to an external news resource, there is a very high possibility that this tweet is closely related to the linked resource. This feature is applied to collect tweets that share the same URL.

- *Hashtag-based selection:*

The # symbol, called hashtag, is used to mark a topic in a tweet or to follow conversation. Hashtag is meant to be identifier for conversations that rotate around the same topic. By including hashtag in a message(tweet), users indicate to which conversations their message is related to. We used this feature to collect tweets that share the same hashtags.

- *Time Difference and Publication dates:*

The time difference is highly important feature for detecting tweets linked indirectly to the conversation. We use the time attribute to efficiently remove tweets having a large distance in terms of time compared to conversation root. Date attribute are highly important for detecting conversations. Users tend to post tweets about conversational topic within a short time period. The euclidean distance has been used to calculate how similar two posts publication dates are.

- *Content-based Similarity:*

The criterion content refers to the thematic relevance traditionally calculated by IR systems standards. We compute the textual similarity between each element in t_i , t_j taking the maximum value as the similarity measure between two messages. The similarity between two elements is calculated using the well-known tf-idf cosine similarity, $\text{sim}(t_i, t_j)$.

After calculating the tweets' scores, a tweet sort relative to each conversation will be established during the reduce phase. Then, for each conversation, we suggest to select the five ranked tweet. The final output will be a conversation set, with an identified root for each one, and enriched by five extra tweets.

4 Experimentations and Results

The following experiment has been designed to gather some knowledge on the impact of our results on end-users. For this experiment, we have selected three events and queried our dataset using Google, Twitter search engine and our approach. Then, we have asked a set of assessors to rate the top-10 results of every search task, to compare these approaches.

4.1 Evaluation Metrics

In order to measure the results quality, we use the Normalized Discounted Cumulative Gain (NDCG) at 10 for all the judged events. In addition, we used a second metric which is the Precision at top 10. In the following, we first describe the experimental setting, then we present the results and we provide an interpretation of the data.

4.2 Experimental Settings

The dataset has been obtained by monitoring microblogging system Twitter posts over the period of March-May 2015. In particular, we used a sample of about 313 000 posts consisted of 1TB of data containing trending topic keywords using Twitter's streaming API. Trending topics (the most famous topic on twitter) have been determined directly by Twitter, and we have selected the most frequent ones during the monitoring period. To evaluate our search tasks' results, we have used a set of 100 assessors with three relevance levels, namely highly relevant (value equal to 2), relevant (value equal to 1) or irrelevant (value equal to 0). Every user was informed of three events happened during the sampling period. For each event we performed three searches using:

1. Google.
2. Twitter Search.
3. Our approach.

The evaluators were not aware from which system the results were retrieved. Every user for each search task was presented with three conversations selections, one for each of the previous options with the corresponding top-10 results.

4.3 Results Interpretation

We compare our conversation retrieval approach with the results returned by Google and by Twitter search engine using two metrics namely the P@10 and the NDCG@10. From this comparison, we obtained the values presented in figure 2 and figure 3 where we notice that our approach overcomes the results given by both of Google and Twitter.

The reason of these promising values is the fact that we combine a set of social conversational features and big data environment characteristics to retrieve conversation may have a significant impact on the users' evaluation. Focusing on the three messages selections, we observe that all conversations obtained with our approach receive higher scores with compared to Google and Twitter's selection. According to the free comments of some users and following the analysis of the posts in the three selections we can see that Google and twitter received lower scores not because they contained posts judged as less interesting, but because some posts were considered not relevant with regard to the searched topic.

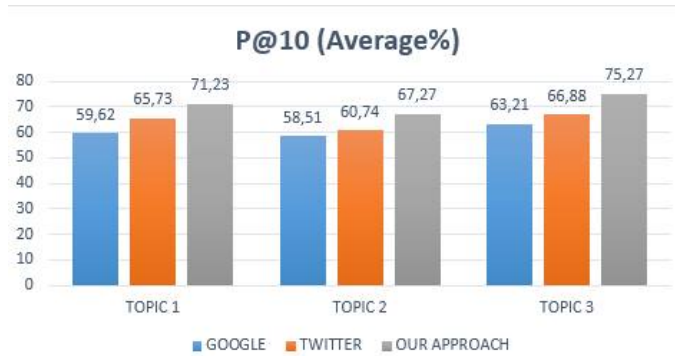


Fig. 2. P@10 Average scores of human evaluation for each approach.

Concentration on the three messages selections we observe that all conversations selections obtained with twitter search has higher scores with respect to Google's selection. These results lead us toward a more general interpretation of the collected data. It appears that the social metrics usage have a significant impact on the users' degree interest in the retrieved posts. In addition, the retrieving conversations process from Social Network differs from traditional Web information retrieval; it involves human communication aspects, like the degree interest in the conversation explicitly or implicitly expressed by the interacting people.

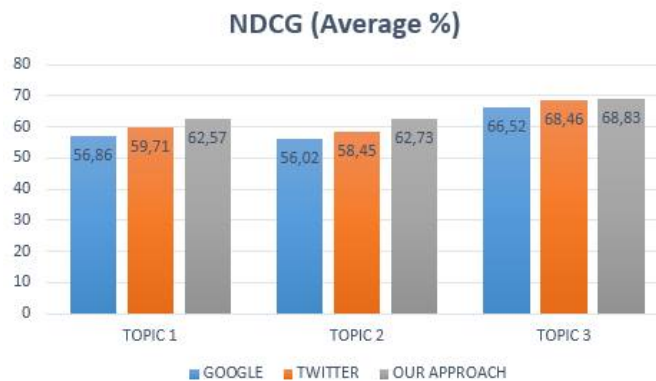


Fig. 3. NDCG Average scores of human evaluation for each approach.

5 Conclusion

Only Big Data applications can enable users to manage social conversations. In addition, Big Data is usually associated to having volume, velocity, variety, variability and complexity which are also features of Social Media Analytics. That is why social media should be explored to make sense of what all the available data means to satisfy users' information needs.

This work proposed a new approach to retrieve conversation on microblogging sites that combine Big Data environment and social media analytics solutions. Our experimental results have highlighted many interesting points. First, combining social media features and the concept of direct conversation on Big data environment improves the relevance and informativeness of information search task and also provides results that are considered more satisfaction with respect to

a traditional information retrieval approach. Future work will further research the conversational aspects by including human communication aspects, like the degree of interest in the conversation and their influence/popularity by gathering data from multiple sources from Social Networks.

References

1. Belkaroui, R., Faiz, R.: Towards events tweet contextualization using social influence model and users conversations. In: Proceedings of the 5th International Conference on Web Intelligence, Mining and Semantics, WIMS 2015, Larnaca, Cyprus, July 13-15, 2015. p. 3 (2015), <http://doi.acm.org/10.1145/2797115.2797134>
2. Bernstein, M., Hong, L., Kairam, S., Chi, H., Suh, B.: A torrent of tweets: Managing information overload in online social streams. In: In Workshop on Microblogging: What and How Can We Learn From It? (CHI '10 (2010)
3. Bollier, D., Firestone, C.M.: The promise and peril of big data. Aspen Institute, Communications and Society Program Washington, DC, USA (2010)
4. boyd, d., Crawford, K.: Six Provocations for Big Data. In: A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society (2011)
5. Bruns, A., Burgess, J.E.: #Ausvotes: how twitter covered the 2010 australian federal election. *Communication, Politics and Culture* 44(2), 37–56 (2011), <http://search.informit.com.au/documentSummary;dn=627330171744964;res=IELHSS>
6. Cha, M., Mislove, A., Gummadi, K.P.: A measurement-driven analysis of information propagation in the flickr social network. In: Proceedings of the 18th International Conference on World Wide Web. pp. 721–730. WWW '09, ACM, New York, NY, USA (2009), <http://doi.acm.org/10.1145/1526709.1526806>
7. Chen, W., Wang, Y., Yang, S.: Efficient influence maximization in social networks. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 199–208. KDD '09, ACM, New York, NY, USA (2009), <http://doi.acm.org/10.1145/1557019.1557047>
8. Cogan, P., Andrews, M., Bradonjic, M., Kennedy, W.S., Sala, A., Tucci, G.: Reconstruction and analysis of twitter conversation graphs. In: Proceedings of the First ACM International Workshop on Hot Topics on Interdisciplinary Social Networks Research. pp. 25–31. HotSocial '12, ACM, New York, NY, USA (2012), <http://doi.acm.org/10.1145/2392622.2392626>
9. Cuzzocrea, A., Song, I.Y., Davis, K.C.: Analytics over large-scale multidimensional data: The big data revolution! In: Proceedings of the ACM 14th International Workshop on Data Warehousing and OLAP. pp. 101–104. DOLAP '11, ACM, New York, NY, USA (2011), <http://doi.acm.org/10.1145/2064676.2064695>
10. Efron, M., Winget, M.: Questions are content: A taxonomy of questions in a microblogging environment. In: Proceedings of the 73rd ASIS&T Annual Meeting on Navigating Streams in an Information Ecosystem - Volume 47. pp. 27:1–27:10. ASIS&T '10, American Society for Information Science, Silver Springs, MD, USA (2010), <http://dl.acm.org/citation.cfm?id=1920331.1920371>
11. Gómez, V., Kappen, H.J., Kaltenbrunner, A.: Modeling the structure and evolution of discussion cascades. In: Proceedings of the 22Nd ACM Conference on Hypertext and Hypermedia. pp. 181–190. HT '11, ACM, New York, NY, USA (2011), <http://doi.acm.org/10.1145/1995966.1995992>
12. Huang, J., Thornton, K.M., Efthimiadis, E.N.: Conversational tagging in twitter. In: Proceedings of the 21st ACM Conference on Hypertext and Hypermedia. pp. 173–178. HT '10, ACM, New York, NY, USA (2010), <http://doi.acm.org/10.1145/1810617.1810647>
13. Jabeur, L.B., Tamine, L., Boughanem, M.: Uprising microblogs: A bayesian network retrieval model for tweet search. In: Proceedings of the 27th Annual ACM Symposium on Applied Computing. pp. 943–948. SAC '12, ACM, New York, NY, USA (2012), <http://doi.acm.org/10.1145/2245276.2245459>
14. Kempe, D., Kleinberg, J., Tardos, E.: Maximizing the spread of influence through a social network. In: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 137–146. KDD '03, ACM, New York, NY, USA (2003), <http://doi.acm.org/10.1145/956750.956769>
15. Kumar, R., Mahdian, M., McGlohon, M.: Dynamics of conversations. In: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 553–562. KDD '10, ACM, New York, NY, USA (2010), <http://doi.acm.org/10.1145/1835804.1835875>
16. Kwak, H., Lee, C., Park, H., Moon, S.: What is twitter, a social network or a news media? In: Proceedings of the 19th International Conference on World Wide Web. pp. 591–600. WWW '10, ACM, New York, NY, USA (2010), <http://doi.acm.org/10.1145/1772690.1772751>
17. Lee, C., Kwak, H., Park, H., Moon, S.: Finding influentials based on the temporal order of information adoption in twitter. In: Proceedings of the 19th International Conference on

- World Wide Web. pp. 1137–1138. WWW '10, ACM, New York, NY, USA (2010), <http://doi.acm.org/10.1145/1772690.1772842>
18. Magnani, M., Montesi, D., Rossi, L.: Information propagation analysis in a social network site. In: *Advances in Social Networks Analysis and Mining (ASONAM), 2010 International Conference on*. pp. 296–300 (Aug 2010)
 19. Magnani, M., Montesi, D., Nunziante, G., Rossi, L.: Conversation retrieval from twitter. In: Clough, P., Foley, C., Gurrin, C., Jones, G., Kraaij, W., Lee, H., Mudoch, V. (eds.) *Advances in Information Retrieval, Lecture Notes in Computer Science*, vol. 6611, pp. 780–783. Springer Berlin Heidelberg (2011), http://dx.doi.org/10.1007/978-3-642-20161-5_93
 20. Magnani, M., Montesi, D., Rossi, L.: Conversation retrieval for microblogging sites. vol. 15, pp. 354–372. Springer Netherlands (2012), <http://dx.doi.org/10.1007/s10791-012-9189-9>
 21. Manovich, L.: Trending: the promises and the challenges of big social data. *Debates in the digital humanities* pp. 460–475 (2011)
 22. Sagioglu, S., Sinanc, D.: Big data: A review. In: *Collaboration Technologies and Systems (CTS), 2013 International Conference on*. pp. 42–47 (2013)
 23. Scellato, S., Mascolo, C., Musolesi, M., Crowcroft, J.: Track globally, deliver locally: Improving content delivery networks by tracking geographic social cascades. In: *Proceedings of the 20th International Conference on World Wide Web*. pp. 457–466. WWW '11, ACM, New York, NY, USA (2011), <http://doi.acm.org/10.1145/1963405.1963471>
 24. Smith, M., Szongott, C., Henne, B., von Voigt, G.: Big data privacy issues in public social media. In: *Digital Ecosystems Technologies (DEST), 2012 6th IEEE International Conference on*. pp. 1–6 (June 2012)
 25. Song, S., Li, Q., Zheng, N.: A spatio-temporal framework for related topic search in microblogging. In: An, A., Lingras, P., Petty, S., Huang, R. (eds.) *Active Media Technology, Lecture Notes in Computer Science*, vol. 6335, pp. 63–73. Springer Berlin Heidelberg (2010), http://dx.doi.org/10.1007/978-3-642-15470-6_8
 26. Teevan, J., Ramage, D., Morris, M.R.: #twittersearch: A comparison of microblog search and web search. In: *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*. pp. 35–44. WSDM '11, ACM, New York, NY, USA (2011), <http://doi.acm.org/10.1145/1935826.1935842>
 27. Wang, D., Wen, Z., Tong, H., Lin, C.Y., Song, C., Barabási, A.L.: Information spreading in context. In: *Proceedings of the 20th International Conference on World Wide Web*. pp. 735–744. WWW '11, ACM, New York, NY, USA (2011), <http://doi.acm.org/10.1145/1963405.1963508>
 28. Yang, J., Leskovec, J.: Patterns of temporal variation in online media. In: *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*. pp. 177–186. WSDM '11, ACM, New York, NY, USA (2011), <http://doi.acm.org/10.1145/1935826.1935863>