What does Everybody Know? Identifying Event-Specific Sources from Social Media

Debanjan Mahata and Nitin Agarwal
Department of Information Science,
University of Arkansas at Little Rock, U.S.A
dxmahata@ualr.edu, nxagarwal@ualr.edu

***Abstract*—Social media is increasingly becoming a popular platform for the public to voice their opinion and present them to a huge audience in the web. The year 2011 saw one of the greatest use of social media in the rise and spread of various events, and has been rightly defined as the year of "Social Media Democracy", with "The Protester" being named as the TIME magazine's person of the year 2011. Due to the power law distribution of the Internet, it is highly likely that the social media sites are buried in the Long Tail. It is therefore, of utmost importance to identify quality social media sources from the Long Tail, for understanding and exploring the real-life events in depth. In this work, we propose a framework to distinguish the disparate sources from social media that provide extremely significant information about various events. Specifically, we propose information theoretic measures to identify "specific" sources for an event (often buried in the Long Tail) and "closer" entities (individuals, groups, organizations, places, etc.) for an event. We also introduce a novel evaluation strategy for validating the proposed measures. Data for the research is collected from various blogging platforms. Experiments demonstrate promising results with interesting findings.**

***Keywords*-event analysis, social media, specificity, closeness, long tail, information gain**

I. INTRODUCTION

In the words of Madeline Albright [11], "Most news networks covered events in Egypt as if they were a football game, when in fact they are more akin to a marathon". Such a statement in context to the recent 'Egyptian Revolution' of 2011, portrays the trade-off between using the mainstream media and the social media as an extensive source of information about various real-life events . Mainstream media often glosses over the intricate details of a particular event and provides summarized coverage of ongoing events from all over the world. On the other hand, we often find a contrasting nature of the social media channels, filled with enthusiastic users. They discuss and write about the events, taking into account even the minor incidents related to them. These genuinely interested users not only follow the events while they occur but also analyze their repercussions, even after they lose their intensity in the mainstream channels. Such a dichotomy between the two media channels motivated us to develop new measures for identifying sources that could potentially offer a 'closer' view of an event in the realm of contemporary social media environment.

Social media has engendered an irreversible change in the way people share information and communicate. Using social media, ideas and perspectives - ranging from innocuous rants to dangerous propaganda - can instantly be broadcasted and shared with the world at large. More recently, social media also became a leading platform for organizing and coordinating different types of events all over the world. In 2011, we saw indigenous uses of social media in mobilizing events such as, 'The Arab Spring', 'Occupy Wall Street', 'Tsunami relief efforts in Japan', 'The London Riots', and the 'The Spanish revolution'. An unprecedented increase in citizen journalistic activities was exemplified. People provided live commentaries for these events and various related incidents on social media sites, encouraging transnational participation. Researchers, journalists, marketers, government bodies, and companies have analyzed such events from the social media sources to study online public opinion [22], diffusion of pandemics [6] and information [4], human behavior [15], socio-political [18] and cultural inferences [2], use of social media during crisis management [21], etc.

Challenges: Despite its utility as an event analysis platform, social media has its own inherent drawbacks, posing difficult challenges in leveraging it. Since the Internet follows the power law distribution [1], enormous population of the buried 'Long Tail' social media sources makes it very challenging to identify quality sources and to collect them. The entities (person, organization, place, etc) mentioned in the sources act as the atomic units of information. Sources which are very 'specific' to an event must contain entities 'closer' or highly relevant to the event. On the other hand, such 'close' entities can be obtained from highly relevant sources. This creates a dilemma in assessing the quality of these sources for event related 'specific' information content, and makes it a nontrivial task. It is also a challenging task to accurately extract the entities from the social media sources, which are mostly unstructured and have colloquial content. Conventional information retrieval based evaluation measures undoubtedly help in identifying the most relevant and authoritative sources, however, these sources may not be the most novel or offer specific information. Therefore, new evaluation measures are also required to estimate the performance of such a work.

Motivation: It is commonly observed that the 'short head' is dominated by the sources from the mainstream media and

978-1-4673-4794-5/12/$31.00 ©2012 IEEE

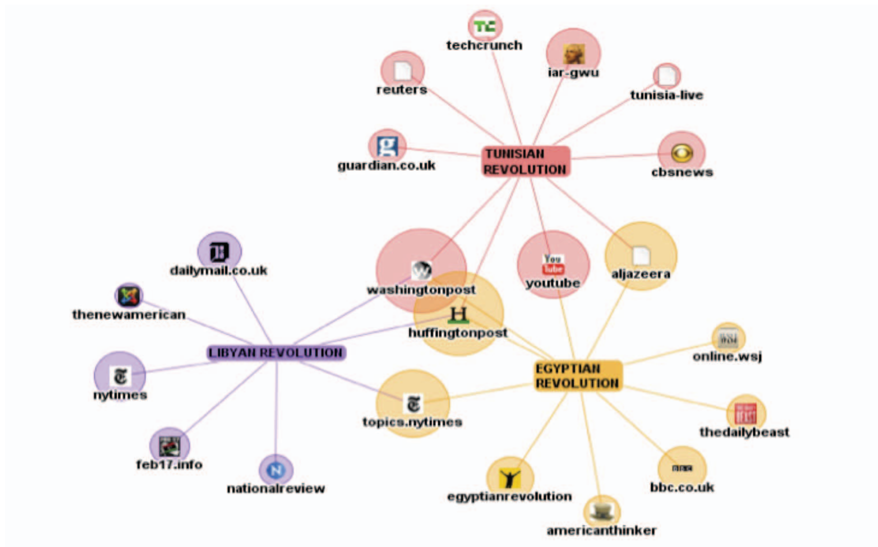

Fig. 1. Top 10 search results for 'Egyptian Revolution', 'Libyan Revolution', and 'Tunisian Revolution' given by Google, visualized using TouchGraph.

the social media sources lie in the 'long tail' [13]. Consequently, the top 10 search results from Google, for 'Egyptian Revolution', 'Libyan Revolution', and 'Tunisian Revolution' are primarily the mainstream media sites instead of the social media sites, as visualized in Fig 1, using TouchGraph [1]. However, in reality the globally popular and local social media channels played an important role in these events [10]. But such results were not visible due to the high ranks assigned to the mainstream media sites. A person who wants to analyze an event may be misdirected by the top results from the popular search engines, which presents the generic mainstream media sites and misses out the more specific sources in social media. Moreover, social media data sources contain unbiased, uninhibited, and unedited opinions from people with diverse backgrounds, against the biased and regulated information in the mainstream media [9]. This motivated us to look for measures and techniques in this paper, that would help us to identify these buried sources providing information highly 'specific' to a concerned event.

Contributions: Considering the challenges circumvallating our research problem, we make major contributions in this paper, towards realizing a framework that can help in identifying such sources from social media and rank them based on their specific information content with respect to the events. We present the framework, along with an objective evaluation strategy to validate our findings. The framework helps in identifying and ranking highly specific sources of information and entities related to the events: 'Egyptian Revolution', 'Libyan Revolution', and 'Tunisian Revolution'. However, the presented work is equally valid for other types of events.

The rest of the paper is organized as follows: Section II reviews the works related to the scope of our problem. Section III defines the problem. Section IV describes the proposed methodology. Section V gives the details of the collected data. Section VI presents the details of the experiments and the evaluation framework. Section VII concludes our work and envisage future directions for our research.

[1] http://touchgraph.com

II. RELATED WORK

In this section, we refer to some of the relevant previous research on analyzing real-life events and identifying quality sources related to them.

User-generated data from various social media platforms, related to real-life events, have been studied to perform wide range of analysis. Socio-political inferences were drawn by studying sentiments and opinions of people towards public and political events from Twitter [19], [22], as well as blogs [2], [18]. Twitter has been extensively used as a source for analyzing information circulated during natural disasters and crisis situations [20], [5]. Tweets related to events have been extracted, summarized and visualized, in order to have a deeper understanding of the events [14]. Our work is different from all such works, and would help in analyzing events from sources and entities, which are highly specific to an event along with the generic ones.

Due to huge number of informal sources generated in social media it is a difficult task to identify high quality sources related to the real-life events. Videos related to the real-life events were identified from the social media sources [12]. Event related contents have been found, leveraging the tagging and location information associated with the photos shared in Flickr [17]. *Becker et al* [3], studied how to identify events and high quality sources related to them from Twitter. In order to identify the genuine sources of information, credibility and trustworthiness of event related information were studied from Twitter [8]. New methods were investigated for filtering and assessing the verity of sources obtained from social media by journalists [7]. The work presented in this paper, finds quality sources related to an event from social media, in terms of the 'specific' information content of the source, and is quiet different from all such works.

III. PROBLEM DEFINITION

The number of sources related to an event in social media is overwhelming. All these sources may not provide useful information and needs to be processed for identifying valuable sources providing specific information about the concerned event. Provided we have a set of events and a set of sources related to each of these events, we need to rank these sources from the most specific to the most generic ones, based on their specific information content.

Given an event E, a set of 'n' sources ϕ_E, we define a function ψ such that:

$$\psi : S_k \to [0, 1] \tag{1}$$

where, $(S_k \in \phi_E)$, is the k^{th} source, so that we can get an ordered set (φ_E) for the same set of sources in ϕ_E, such that:

$$\varphi_E = \{S_1, S_2, ..., S_i, S_j, ..., S_n \mid \psi(S_i) \geq \psi(S_j), i < j\} \tag{2}$$

IV. METHODOLOGY

We develop a framework to rank sources, from the most specific to the most generic ones in terms of the specific information content related to an event. In order to rank them we need to identify the basic units of information in these sources. Real-life events are mostly associated with unique entities (person, place, organizations, etc), which characterize it and are highly related, or 'closer' to it, distinguishing them from the other events. At the same time there are entities, which are generic. So, if a source contains more 'closer' entities for an event, then there is a higher chance for that source to be more 'specific'. The formulations of 'specificity' (κ) of a source, which helps in identifying the 'specific' sources and 'closeness' (τ) of entities are explained in details, in Section IV-A and Section IV-B , respectively. The 'closeness' (τ) measure of entities are further used for constructing the event dictionaries (Section IV-C).

A. Specificity

In order to select the highly specific sources, we propose a novel 'specificity' measure, which estimates the unique information that a source S_k can offer vis-$\grave{a}$-vis an event E_i. It is important to note here that a source's specificity is always estimated with respect to a given event. The measure draws upon the theory of information gain and is defined as κ. Mathematically,

$$\kappa = IG(E_i, S_k) = H(E_i) - H(E_i, S_k) \tag{3}$$

where $IG(E_i, S_k)$ denotes the information gain for a source $(S_k \in \phi_{E_i})$ related to an event $(E_i \in \xi)$, $H(E_i)$ is the total entropy for the event E_i, and $H(E_i, S_k)$ is the total entropy of the event E_i given the source S_k. ϕ_{E_i} denotes the set of sources for the event E_i. ξ denotes the set of events selected for the study. S_k is the k^{th} source belonging to ϕ_{E_i}. E_i is the i^{th} event belonging to ξ. Since, $H(E_i)$ is constant for every event,

$$IG(E_i, S_k) \propto -H(E_i, S_k) \tag{4}$$

So we only calculate the values for $H(E_i, S_k)$ in order to find κ. After finding κ for each source $(S_k \in \phi_{E_i})$, we rank the sources based on the value of κ and arrange them in decreasing order from the most specific to the most generic ones. The formulation of κ discussed above is generic and can be implemented in various ways. Section VI-B explains a specific implementation used in this work.

B. Closeness

We find a measure of closeness τ for entities related to an event $(E_i \in \xi)$. Such a measure highlights the significance of an entity '*e*' with respect to an event E_i. We formulate it by using the concept of conditional probability. Mathematically, the measure is given as:

$$\tau = P(e, E_i) = P(e|E_i) * P(E_i) \tag{5}$$

where $P(e, E_i)$ is the probability of an entity '*e*' to be associated with an event $(E_i \in \xi)$. $P(e|E_i)$ is the probability of the occurrence of an entity '*e*', provided that the event E_i has occurred and $P(E_i)$ is the probability of the occurrence of the event E_i. We calculate τ for each entity and arrange them according to their decreasing order of closeness towards an event E_i. The formulation of τ discussed above is generic and can be implemented in various ways. Sections IV-C and VI-A explains a specific implementation used in this work.

C. Event Dictionary

It is an evolving dictionary of all the entities along with their 'closeness' scores for an event. In order to find highly specific sources we need to find closer entities and in order to find closer entities we need to have a set of highly specific sources. We solve such a problem by constructing the event dictionaries using sources, that are completely separate from the ones used for the experiment. These sources act as the seed sources for initializing the event dictionary, which is later on updated with the help of the specific sources identified by our framework. We calculate the 'closeness' scores for the entities in order to construct the event dictionary, by using the following formula based on tf_idf measure [16] from information retrieval literature:

$$\tau = P(e, E_i) = ef\_iE_if = ef(e, E_i) * iE_if(e) \tag{6}$$

where,

$$iE_if(e) = log(\frac{\mid \xi \mid}{\mid E_i : e \in E_i \mid}) \tag{7}$$

$E_i \in \xi$, and 'e' is an entity extracted from the set of sources. The term $ef(e, E_i)$ denotes the frequency of occurrence of the entity 'e' in the set of sources for the the event E_i. The term $iE_if(e)$ denotes the inverse event frequency for the entity 'e', which is obtained by dividing the total number of events $\mid \xi \mid$, by the total number of events in ξ that mentions the entity 'e' $(\mid E_i : e \in E_i \mid)$.

TABLE II
TOP 5 ENTITIES IN THE EVENT SPECIFIC AND THE GLOBAL DICTIONARIES CONSTRUCTED FOR THE SET OF EVENTS.

Egyptian Revolution Specific Dictionary	Tahrir Square, Egyptian government, Gigi Ibrahim, Alexandria, Wael Abbas.
Libyan Revolution Specific Dictionary	Tripoli, Muammar Al Gaddafi, North Atlantic Treaty Organization, Chad, United Kingdom
Tunisian Revolution Specific Dictionary	Tunisian government, Lin Ben Mhenni, Samir Feriani, Kasbah Square, RCD
Socio-Political (global) event dictionary	Twitter, Iranian Government, Tear gas devices, Facebook, Big Social network

TABLE I
DETAILS OF DATA COLLECTED.

Service Used	Event	Number of Posts
GlobalVoices	Egyptian Revolution	234
	Libyan Revolution	86
	Tunisian Revolution	77
Google Blogger	Egyptian Revolution	579
	Libyan Revolution	600
	Tunisian Revolution	484
Icerocket Blog Search	Egyptian Revolution	5900
	Libyan Revolution	2198
	Tunisian Revolution	1220

V. DATA COLLECTION

Data for the study are collected from GlobalVoices[2], Blogger[3] and Icerocket[4], respectively. The details of these datasets are given in Table I. The data from GlobalVoices is used for constructing the event dictionaries. We choose GlobalVoices as it is a portal where bloggers and translators work together to make reports of various real-life events, from blogs and citizen media everywhere. These blogs are manually curated and translated into English from their native language. We collect blog posts related to the three events from Blogger using Google Search, and from various other platforms using Icerocket blog search. The collected blog posts are parsed for extracting URL (blog, blogpost), blog text, entities, and its rank in the respective search engine used for collecting it. We use AlchemyAPI[5] in order to extract entities. These datasets will be made available on request.

VI. EXPERIMENT

Experiments are performed for constructing the event dictionaries using GlobalVoices dataset and identifying the 'specific' sources from the social media using the Blogger, and Icerocket datasets . We introduce a novel objective evaluation framework in order to validate our findings.

A. Constructing Event Dictionaries

Following steps are taken to construct the event dictionaries:

1) Entities are extracted from all the sources using AlchemyAPI and their τ values are calculated using the formula mentioned in Section IV-C.
2) Entities are then ranked according to the decreasing τ values.
3) For each individual event, the event specific dictionaries are constructed by selecting the entities having positive values of τ, as these are the entities, which have the highest probabilities to be associated with the selected event.
4) The entities having zero or negative values of τ are used for constructing the global dictionary, as they have the lowest probabilities to be associated with any particular event. These are the generic entities and are common across the set of events under study.
5) After constructing the dictionaries the values of τ are scaled between 0 and 1.

Table II shows the top five entities in the event specific dictionaries for each of the events and a socio-political global dictionary for the set of events. The entities in the event specific dictionaries, when present in a source makes it highly 'specific' to that particular event and contributes in gaining information about it. On the other hand, the entities of the global dictionary provides shallow information about a specific event, and are useful in learning about a category of events, in this case, socio-political uprisings in the middle east.

B. Identifying 'Specific' Long Tail Social Media Sources

The 'specificity' (κ) score of a source related to an event ($E_i \in \xi$) is calculated as follows:

1) Each source ($S_k \in \phi_{Ei}$) is represented by an entity vector $\varepsilon_{E_i,S_k} =< e_1, e_2, e_3, ..., e_n >$, created from the entities extracted from it. The sources having entities less than three are ignored, as they don't contribute much in knowing about the event. The value was decided after carefully studying the distribution of the number of entities in the sources.
2) The entities in the entity vector of a source are assigned their corresponding τ values from the related event dictionary.
3) κ is calculated for each source ($S_k \in \phi_{Ei}$) using the following formula:
$$\kappa = -H(E_i, S_k) = \frac{\sum_{i=1}^{n} f_i \tau_i}{\sum_{i=1}^{n} f_i} \tag{8}$$
where, 'n' is the number of entities in the source S_k, f_i is the frequency of occurrence of the i^{th} entity in the source S_k, and τ_i is the 'closeness' value for the i^{th} entity with respect to the event E_i.
4) Sources are then ranked according to their decreasing κ value.

We further analyze the ranking of the top K specific sources as identified by our framework and observe the difference in their rankings as assigned by the search engines and as assigned by our framework. Fig 2 shows ranks assigned by Google and Icerocket for the top 10 sources for each event,

[2]http://globalvoicesonline.org
[3]http://blogger.com
[4]http://icerocket.com
[5]http://alchemyapi.com

Egyptian Revolution

Specificity Based Ranking	Google Search Ranking
1	59
2	286
3	400
4	277
5	55
6	202
7	6
8	9
9	313
10	374

Libyan Revolution

Specificity Based Ranking	Google Search Ranking
1	13
2	329
3	9
4	194
5	24
6	311
7	364
8	204
9	374
10	184

Tunisian Revolution

Specificity Based Ranking	Google Search Ranking
1	162
2	40
3	420
4	459
5	72
6	181
7	152
8	440
9	99
10	174

Egyptian Revolution

Specificity Based Ranking	Icerocket Blog Search Ranking
1	75216
2	10607
3	53924
4	56604
5	9831
6	25790
7	1
8	99925
9	94614
10	53924

Libyan Revolution

Specificity Based Ranking	Icerocket Blog Search Ranking
1	47276
2	11751
3	4900
4	22501
5	4
6	11040
7	43520
8	11751
9	41631
10	18271

Tunisian Revolution

Specificity Based Ranking	Icerocket Blog Search Ranking
1	9713
2	42985
3	36335
4	3843
5	46784
6	42645
7	99
8	1
9	63141
10	42645

Fig. 2. Rankings of the sources from Google Blogger and Icerocket based on 'specificity' (κ), and the rankings assigned by Google Search and Icerocket.

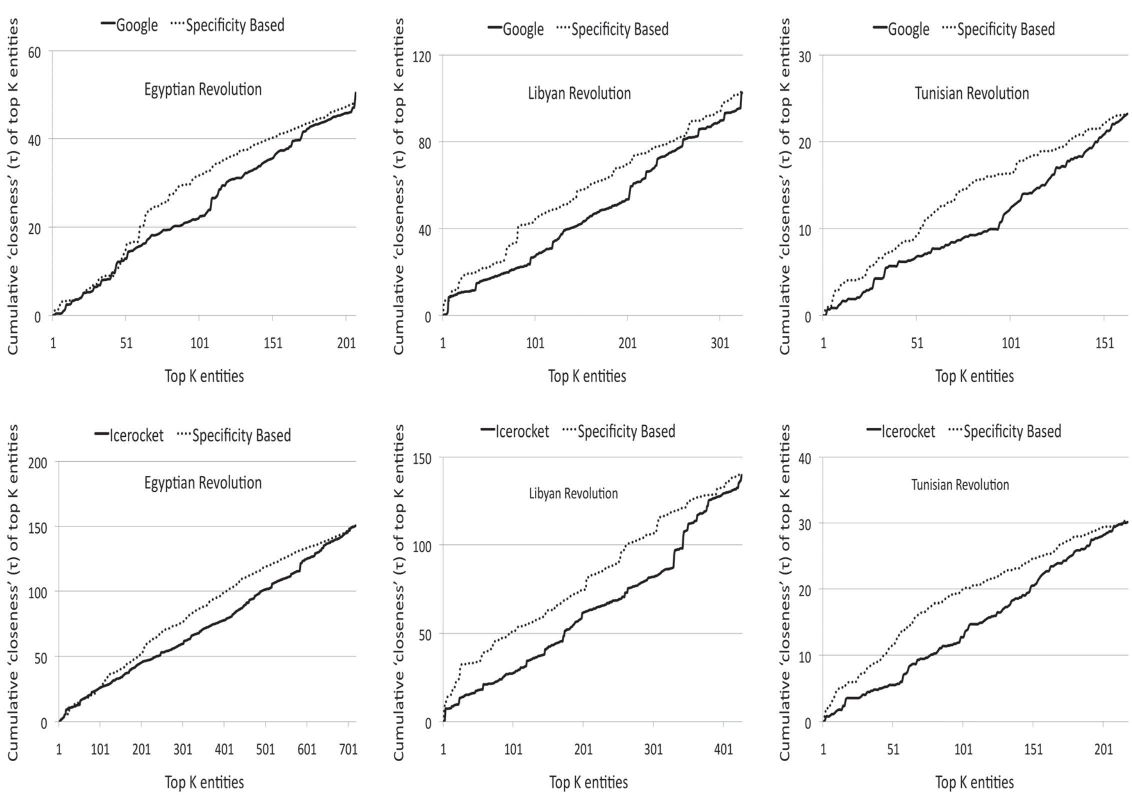

Fig. 3. Validating 'specificity' (κ) of sources from Google Blogger and Icerocket.

ranked according to 'specificty' (κ) value. We conclude, that our framework could identify the sources, which often gets buried in the Long Tail and has the potential for presenting valuable information about the event.

C. Validating 'Specificity' Of The Long Tail Sources

The results obtained in the previous section shows that there are many sources in the Long Tail which contain 'specific' information about an event, yet they get buried. Next, we propose a novel strategy to demonstrate the effectiveness of the 'specificity' (κ) measure in identifying highly informative sources.

Due to lack of benchmark datasets we use the search results obtained from Google and Icerocket Blog Search as baselines for validation. Search engines are designed to give the most relevant sources, also containing the most relevant and 'close' entities, related to a query for a given event. As these 'close' entities have a very high probability to be associated with the event, we expect to gain valuable information about the event from these entities making the highly ranked sources very 'specific' to the event. We use this notion to propose a novel evaluation strategy, showing that when sources related to an event are ranked according to 'specificity' (κ), they provide more valuable information than the ranking order given by the search engines for the same set of sources.

We compare 'closeness' (τ) values of the entities obtained from the sources ranked according to the search engines and 'specificity' (κ) values. From the two differently ranked lists, each source is visited and the entities are extracted from them. The 'τ' values are assigned to these entities by referring to the respective event specific dictionary and they are ranked in descending order of their 'τ' values. As we traverse the two list of sources, we obtain two list of same entities, arranged in different orders depending upon the ranking of the sources. In order to show the comparison in the gain of information from the two lists, we take the top 'K' entities from each list and calculate the sum of their 'τ' values and plot them against the value of 'K' in Fig 3. We start from K=1 and go on increasing its value till the number of entities are exhausted in both the lists. It is evident from Fig 3, that the curves based on 'specificity' (κ) quickly gains over the curves based on the search engines. This shows that information about the event is gained quicker using our framework, which could identify 'specific' sources earlier than the search engines. This in turn implies that the sources ranked higher by our framework are more 'specific' than the ones ranked by the search engines. As we already observed that these sources are often Long Tail sources, we can conclude that Long Tail sources that are ranked lower by the search engines, when identified are often more 'specific' than the highly ranked short head sources. When presented earlier, they help in learning useful information about the event due to the presence of 'closer' entities in them.

VII. Future Work and Conclusion

In this paper, we proposed and validated a framework that is capable of identifying sources buried in the Long Tail and provide novel and valuable information specific to an event from social media. Towards this direction, we developed measures for finding 'specific' sources and 'closer' entities related to real-life events. We delineated overarching challenges that underlie the event analysis research area from the social media lens. We perceive the proposed framework as the first step towards addressing these challenges, including overabundance of Long Tail social media sources, and above all, evaluation in the absence of ground truth. We proposed a novel validation framework, leveraging most widely used search engines as baselines. It was observed that the search engines ranked the 'specific' sources quite low in the returned results, thereby reducing the chances of their discovery.

In future, we plan to explore the utility of the proposed framework in studying various socio-technical, demographic and behavioral patterns. Social media would continue to grow as a primary source for analyzing real-life events and there is a need for developing tools and techniques for analyzing the events through the lens of social media. We are interested in making our framework more robust and dynamic so that it can cater to the challenges of analyzing all types of real-life events from the sources available in various social media platforms.

Acknowledgment

The research is supported in part by grants from the US Office of Naval Research (Award: N000141010091) and the US National Science Foundation (Awards: IIS-1110868, IIS-1110649). The support is gratefully acknowledged. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agencies.

References

[1] L. Adamic and B. Huberman. Power-law distribution of the world wide web. *Science*, 287(5461):2115–2115, 2000.

[2] N. Agarwal et al. Understanding group interaction in blogosphere: A case study. In *2nd International Conference on Computational Cultural Dynamics (ICCCD08), Washington DC*, 2008.

[3] H. Becker et al. Selecting quality twitter content for events. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media (ICWSM11)*, 2011.

[4] Borge-Holthoefer et al. Structural and dynamical patterns on online social networks: The spanish may 15th movement as a case study. *PLoS ONE*, 6(8):e23883, 08 2011.

[5] C. Cheong. Social media data mining: A social network analysis of tweets during the australian 2010-2011 floods.

[6] C. Chew and G. Eysenbach. Pandemics in the age of twitter: Content analysis of tweets during the 2009 h1n1 outbreak. *PLoS ONE*, 5(11):e14118, 11 2010.

[7] N. Diakopoulos et al. Finding and assessing social media information sources in the context of journalism. In *Proc. Conference on Human Factors in Computing Systems (CHI)*, 2012.

[8] M. Gupta et al. Evaluating event credibility on twitter. 2012.

[9] N. Hamdy and E. Gomaa. Framing the egyptian uprising in arabic language newspapers and social media. *Journal of Communication*, 2012.

[10] Z. Harb. Arab revolutions and the social media effect. *M/C Journal*, 14(2), 2011.

[11] R. Jeffrey Abramson. *Networks and Citizenship: Using Technology for Civic Innovation*. The Aspen Institute, 2012.

[12] L. Kennedy and M. Naaman. Less talk, more rock: automated organization of community-contributed collections of concert videos. In *Proceedings of the 18th international conference on World wide web*, WWW '09, pages 311–320, New York, NY, USA, 2009. ACM.

[13] LOmariba. Is new media posing a serious challenge to traditional media?". Technical report, University of Westminster, 2009.

[14] A. Marcus and others. Twitinfo: aggregating and visualizing microblogs for event exploration. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, CHI '11, pages 227–236, New York, NY, USA, 2011. ACM.

[15] M. Mendoza, B. Poblete, and C. Castillo. Twitter under crisis: can we trust what we rt? In *Proceedings of the First Workshop on Social Media Analytics*, SOMA '10, pages 71–79, New York, NY, USA, 2010. ACM.

[16] J. Ramos. Using tf-idf to determine word relevance in document queries. In *Proceedings of the First Instructional Conference on Machine Learning*, 2003.

[17] T. Rattenbury, N. Good, and M. Naaman. Towards automatic extraction of event and place semantics from flickr tags. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '07, pages 103–110, New York, NY, USA, 2007. ACM.

[18] V. Singh et al. Mining the blogosphere from a socio-political perspective. In *Computer Information Systems and Industrial Management Applications (CISIM), 2010 International Conference on*, pages 365 –370, oct. 2010.

[19] M. Thelwall et al. Sentiment in twitter events. *Journal of the American Society for Information Science and Technology*, 62(2):406–418, 2011.

[20] S. Vieweg et al. Microblogging during two natural hazards events: what twitter may contribute to situational awareness. In *Proceedings of the 28th international conference on Human factors in computing systems*, CHI '10, pages 1079–1088, New York, NY, USA, 2010. ACM.

[21] D. Yates and S. Paquette. Emergency knowledge management and social media technologies: A case study of the 2010 haitian earthquake. *International Journal of Information Management*, 31(1):6 – 13, 2011.

[22] A. Younus et al. What do the average twitterers say: A twitter model for public opinion analysis in the face of major political events. *Social Network Analysis and Mining, International Conference on Advances in*, 0:618–623, 2011.