Monitoring Collective Attention During Disasters

Xingsheng He*†, Yu-Ru Lin†
*Department of Electronic Science and Technology, University of Science and Technology of China
†School of Information Sciences, University of Pittsburgh
Email: hxs@mail.ustc.edu.cn, yurulin@pitt.edu

***Abstract*—The proliferation of vast information shared through social media and other communication technologies has led us to an era of attention scarcity. Effective crisis response requires both a systematic understanding of how collective attention emerges during disaster events and robust techniques for monitoring the public's attention shift due to the events.**

However, "attention" is an abstract concept that is very challenging to characterize. In this work, we propose an *attention shift network* to systematically analyze the dynamics of collective attention in response to real-world exogenous shocks such as disasters. Using hashtags appeared in Twitter users' complete timelines around a violent terrorist attack – the Paris attacks occurred on November 13, 2015 – we thoroughly investigate the properties of network structures and reveal the temporal dynamics of the collective attention under both regular time and exogenous shock.

I. Introduction

The advent of social media has provided an unprecedented way for people to easily get access to various information at almost no cost, moreover, the rise of social media also facilitates researchers to study the collective human behavior at very large scales. In recent years, there has been growing interest in the use of social media in crisis response [1], such as natural disasters [2], terrorist attacks [3] and political riots [4]. Effective crisis response such as in time warning/evacuation is a vital part of crisis management, which may be beneficial for disaster-affected communities and professional responders, thus, understanding the mechanisms and the dynamic patterns of collective behavior during disasters becomes more and more important.

In this paper, we are interested in quantitatively capturing users' attention under exogenous shocks on Twitter. Anticipatedly, online social media users exhibit distinct behavior patterns under disasters comparing to normal state – typically their "focus points" are usually scattered over a variety of concerns, but it can suddenly concentrate on specific events and then gradually disperses elsewhere – which we refer to as "collective attention". The vast digital traces behind social media create invaluable opportunities to explore the distinct patterns of collective attention, but the rapid time-variant nature and large volumes of such data make it very challenging for many reasons [5]. First, "attention" is an abstract concept that is difficult to characterize, many previous works have defined and captured "attention" from various aspects. Lehmann et al. [6] study different types of collective attention by analyzing the temporal hashtag adoption patterns. Sasahara et al. [5] detect the burst-like increase in tweet numbers and semantic terms as a sign of emerging collective attention. Other work [7] utilizes the trending topics provided by Twitter as a representation of current public focused attention and try to predict future trending list. Wang et al. [8] employ users' clicks stream on web forms as a proxy of collective attention. However, few works have studied the collective attention changing patterns under disasters. Second, collective attention shift process is driven by the interchange between the disordered and synchronized states of multiple individuals, how to capture the individual user's attention on social media is the basis for analyzing attention in the collective level but has not been explored. Moreover, the overwhelming amount of tweet conversations flooded by numerous trivial messages during disasters makes meaningful analysis very difficult – diving into the concrete context of each piece of information is essential for accurate analysis but also will make it overwhelmingly complicated.

In this paper, by using a large corpus of geo-tagged tweets around a shocking terrorist attack in 2015 – the Paris attacks[1] on November 13 – we employ hashtags – the representative eccentric topics for the user-generated content – to delegate the current attention of each individual. We extract the hashtag adoption streams from a user's tweet timeline as the trace of his/her attention shift process. Additionally, when we consider the similar process for a group of users, we could get a snapshot of the collective attention shifting process which would possibly help to discover on-going topics and trace the topic evolutions.

We examine how collective attention on social media systematically shifts in response to real world disaster and the main contributions are as follows: First, we propose a novel way to represent the collective attention shifting process. Second, based on the proposed *attention shift network*, we show that a set of metrics exhibit great potential to detect or quantitatively define the severe level of disasters. Third, we reveal how the network changes during a disaster.

II. Related Work

In the context of social media, many authors utilized the vast digital data to study different aspects of collective attention. Wu and Huberman [9] modeled the decay in news popularity but the underlying mechanism is not explained explicitly. Sasahara [5] leveraged the change in tweets volume and semantic terms to detect the emergence of collective attention. Lee [10]

[1] https://en.wikipedia.org/wiki/November_2015_Paris_attacks

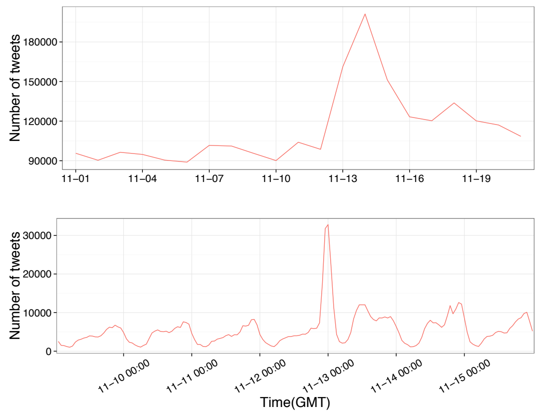


Fig. 1. The top shows the number of tweets per day from November 1st to 22nd and the bottom is the number of tweets per hour from 11-10 00:00 to 11-16 23:00 for Paris users.

examined the problem of collective attention spam when users collectively focus around a phenomenon and found unlike other spam types, collective attention spam relies on the users themselves to seek out the content. Lin et al. [11] studied the conditions of shared attention as many users simultaneously tune in with the dual screens of broadcast and social media to view and participate, they demonstrated how bursts of activity on Twitter during media events significantly alter underlying social processes of interpersonal communication and social interaction. Wang et al. [8] built a attention network to study the popularity of new and old forum threads and found the network preserves self-similarity qualified by a new form of Zipf's law. Many others focused on predicting future trending topics [7], classifying different types of collective attention [6] and how information spread in social networks under the constraint of attention [12].

In terms of disasters, it's difficult for emergency management to monitor collective human behaviors [13]. There are analyses from both empirically study [14] and modeling [15] methods to tackle the collective human behavior under extraordinary events. Wen et al. [16] studied the emotional changes of local Paris users after the Paris attacks. But in this paper, we systematically scrutinize the collective attention under disasters, putting representing and monitoring this abstract concept into a framework.

III. Characterizing Collective Attention

A. Data Collection

We collect Twitter data related to a violent terrorist attack in 2015: The Paris attack on November 13th. Our data collection needs to track the collective attention change across multiple time points: before, during and after the attack. To minimize the selection bias, we primarily study the collective attention of local users. The data collection process is built as follows: First, to identify a group of Paris users, we obtain a set of users who posted geo-tagged tweets from Paris area with a time window of four weeks after the attacks through Twitter API. Then, for each user, we trace back his/her historical tweets through Twitter REST API and retain the users who have also posted tweets before the attack. Finally, the data set contains 16950 users and 1.74 million tweets posted by these users from Oct 31st to Nov 20th. Fig. 1 shows the volume of daily tweets posted by the retained geocodable users starting from Nov 1st and the volume of hourly tweets in the event week. Clearly, there is a burst in the tweet volume on the day of the attack.

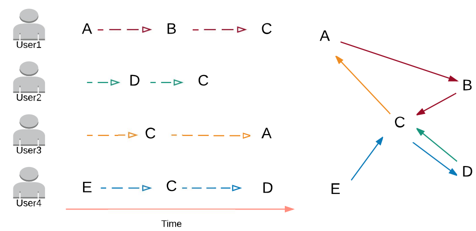


Fig. 2. An example of users' hashtag usage stream and the corresponding collective attention shift network. The left shows given a time interval, the sorted hashtags extracted from the tweet streams for each of the anonymized users. In the right part, the nodes are hashtags and the directed, weighted links represent the consecutive hashtag adoptions, i.e., users' switch to use different hashtags in their consecutive tweets.

B. Collective Attention Shift Model

Fig. 2 presents an example of collective attention shift network for twitter users. In this paper, we study the attention network on the hourly basis. We first chronologically sort the entire dataset and divide them into hourly pieces. After that, for each user, we extract all successive pairs of hashtags from his/her tweets and add corresponding edges in the collective attention network. After iterating over all the users who had generated tweets in that period, we could get a directed, weighted network representing the hashtag shifting situation in the collective level. The constructed collective attention shift network can be deemed as the snapshots of attention competition between different topics/events.

It is noteworthy that hashtags are not necessarily to appear in each of the tweets. Due to the limit of user set size, we may confront the problem of hashtag scarcity especially in the midnight hours, because few tweets in the late night leads to fewer hahstag adoptions. To deal with this problem, when building the attention network at time t, we use four hours' tweets stream prior to t to increase the tweets volume. For example, the attention network at 11:00 pm is built upon tweets from 7:00 pm to 11:00 pm. Another advantage of using multiple hours' tweets stream for hourly analysis is that two ensuing attention network are based on a considerate portion of common data, resulting in the effect of smoothing when we explore the temporal changes over the network characteristics and we will discuss the metrics later in detail. Also, in the following sections, we use "collective attention network", "attention shift network", "attention network" alternatively to refer to as the same concept.

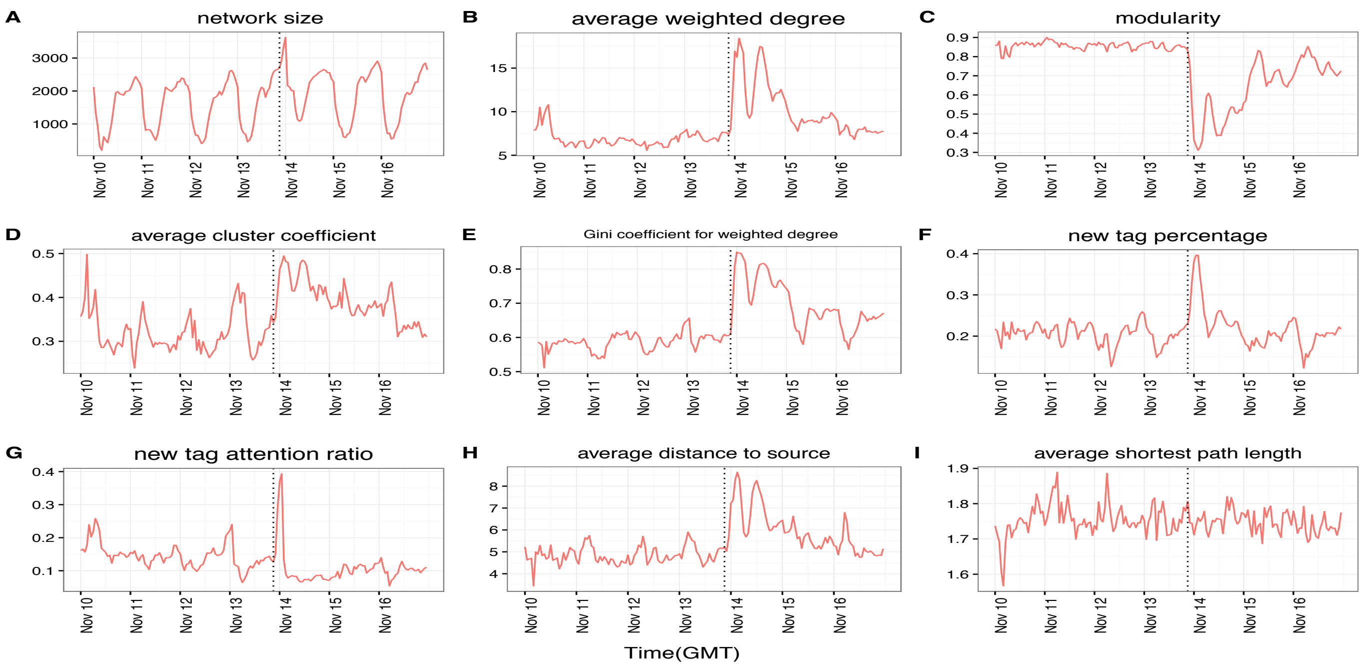


Fig. 3. The representative metrics of the hourly attention shift network over time, starting from 11-10 00:00. The black dashed line in each figure denotes the Paris attacking starting time.

C. Metrics For Network Characteristics

To thoroughly depict the network structure and capture the collective attention shifting patterns under exogenous shock, we list the following metrics we used in this paper and its corresponding functionality.

Network size is the number of hashtags in the graph. It's a simple yet straightforward metric that could mirror the networks' shrink and expansion over time.

Average weighted degree reflects the attention shift speed. Increasing in tweets volume or more hashtags adoption than usual will lead to a higher value.

Modularity inspects the connection between different hashtag classes. A high *modularity* value means the diverse hashtag types as well as the decentralized collective attention [17].

Average clustering coefficient perceives whether the collective attention is likely to shift at a local scale [18]. A higher value will be obtained if most users pay close attention to the same topic and use similar hashtags.

Gini Coefficient for weighted degree measures the level of degree concentration [11], denoting whether a few hashtags have become dominant at that time and Gini Coefficient equals 1 representing perfectly concentrated attention. Like the power-law exponent, Gini coefficient can be used to measure the preferential patterns but in a more general way.

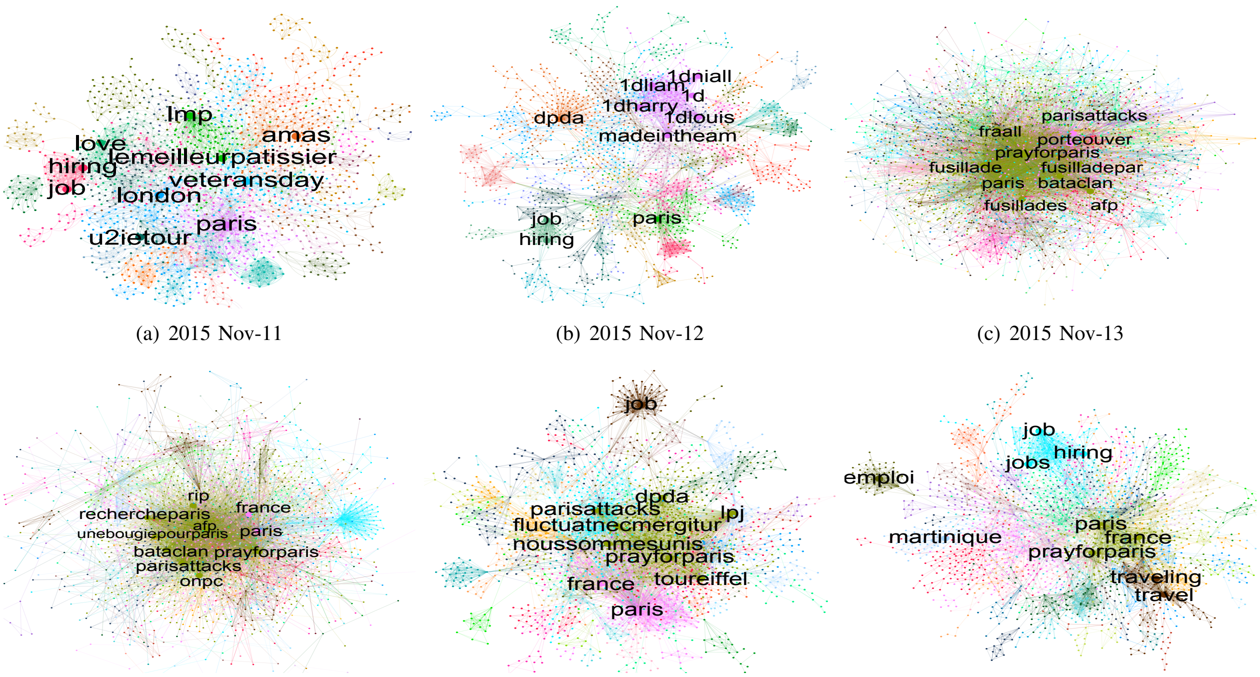
New tag percentage is the number of newly emerged hashtags over the network size. By "new", we mean for a particular set of users, when we build the attention shift network starting at timestamp t, we also check if a hashtag h had been used by these users between $[t-t_0, t]$. If the hashtag h had been used some time prior to the starting timestamp t, we say h is an "old-fashioned" hashtag, otherwise, h is a "new" hashtag. As an exogenous shock bursts and goes viral in the virtual space, lots of brand new hashtags will be invented to aggregate the related information or to express personal opinions [19]. To some extent, the number of new tags could reveal the level of the exogenous shock. In this paper, t_0 is set to a week, i.e., 168 hours.

New tag attention ratio is the percentage of overall attention imposed on newly invented hashtags. We use H and T to represent the whole hashtag set and the "new" hashtag set in an attention shift network, respectively. The *new tag attention ratio* is defined as :

$$r = \frac{\sum_{j \in T} k_{in}^{j}}{\sum_{i \in H} k_{in}^{i}}, \tag{1}$$

where k_{in}^{i} is the weighted in-degree of node i.

Average distance to source. First we add two artificial nodes: "source" and "sink" to the collective attention network which makes both on the network level and on the node level satisfy "flow conservation". Concretely, for each of the nodes in the network whose weighted in-degree is smaller than its weighted out-degree, we will add a weighted edge from "source" to this node, balancing this node such that its in-flow (weighted in-degree) is identical to out-flow (weighted out-degree). Similarly, for the nodes have a larger weighted in-degree, we will add a weighted edge from this node to "sink". After this additional step, in the constructed networks, all the nodes in the network are balanced (weighted in-degree equals weighted out-degree) except the "source" and the "sink". The distance to source d_i as the average steps from the "source"



(a) 2015 Nov-11 (b) 2015 Nov-12 (c) 2015 Nov-13

(d) 2015 Nov-14 (e) 2015 Nov-16 (f) 2015 Nov-19

Fig. 4. Temporal evolution of the collective attention shift network before, during and after the Paris attacks. Each network is built by using the tweets between 20:00:00 and 23:59:59 GMT of each day. The Paris terrorist attacks began at 20:20 GMT on 11/13/2015. Before the event, the network has a clear community structure, reflecting the scattered attention among various topics. After the event happened, a few nodes becomes the hubs that attract most of the attention. With time passing by, the collective attention shift network gradually recovers to the normal state. In each network, we show the top ten hashtags based on weighted in-degree.

node to the ith hashtag, which can be calculated using the Markov property of the attention shift network [8]. *average distance to source* is defined as the average steps from the "source" node to each hashtag.

Average shortest path. Considering the attention shift network is a directed, weighted network, usually there won't be a directed path for each of the pairwise hashtags in the network. Unlike the conventional concepts, after we add the "source" node, we define the shortest path length s_i as the shortest hops from "source" to hashtag i, then we take the average over all the hashtags.

Note that among the aforementioned metrics, *average distance to source* and *average shortest path* will utilize the artificial nodes "source" and "sink", the rest of them depend on the original attention shift network as illustrated in Fig. 2. For the sake of simplicity, when calculating *average clustering coefficient* and *modularity*, we exclude the edge weight information.

IV. Attention Shift Under Exogenous Shocks

Intuitively, tweet activity should obey a regular circadian rhythm in normal situations according to the human sleep/wake up patterns. Also, the collective attention may be imposed on various topics in regular time due to the exuberant contents of social media. When a mega-disaster come to real life, the tweet stream volume tend to have a burst-like increase as well as the concentrated attention on the related event. The attention shift network should be able to well reflect the above phenomenon and help to track the stealthy changes in collective attention under different contexts.

In Fig. 3, we can observe the periodic pattern on *network size* except for the several hours immediately after the Paris attacks which precisely reflects the burst in tweet volume and the existence of exogenous shock. Meanwhile, the percentage of newly emerged hashtags clearly outpaces the normal level which coincides with the analysis in [19] that new hashtags are more likely to emerge after influential social events. We also can observe a similar pattern on *new tag attention ratio* that newly emerged hashtags attracts a lot more attention than usual right after the disaster.

Besides detecting the shock, *average cluster coefficient* together with *average weighted degree* also describe the viewing patterns of users. We find that *average cluster coefficient* only achieves a higher value at midnight hours when the network size is fairly small. But several hours after the attack, when *network size* reaches its maximal value, the *average cluster coefficient* reaches a high level and then slightly decreases over time. A higher *average cluster coefficient* implies the collective attention is more likely to transfer at a local scale, denoting the collective attention is shifting through similar topics. With the sharp rise in *average weighted degree* which embodies the shifting speed, we can come up with an picture that the Paris users are anxiously viewing the latest terrorist attack messages from different aspects(the casualties, the social support activity, the attack locations, etc) one by one, trying to get as many information as possible. However, *average shortest path length* fluctuates heavily and gives little hints of the network changes, partially because in the attention network, many nodes already connect to the "source" node, the small-world effect maintains *average shortest path* at a very small value.

Collective attention shift network is the snapshot of attention competition between different topics. Notably, the attention shift network has an exceptionally high yet consistent modularity value around 0.8 before the Paris attacks. Such a high modularity clearly indicates the distinct community structure together with the balanced distribution of users' attention among diversified topics. In Fig. 4, we visualize the attention network at $00:00$ am on some selected days and label the top 10 popular hashtags based on weighted in-degree. In Fig. 4(a) and (b), In Fig. 3, after the attack happened, we notice that the clear community structure is broken a sharp rise in the *Gini coeffcient for weighted degree* in Fig. 3(E), which denotes a few nodes gain the most attention from the public, usually representing some extreme exogenous shocks happened and been fiercely discussed on the social media. From Fig. 4(c) and (d) , the most popular hashtags are also tightly connected, once again a sign of topic monopoly after remarkably attentive social events. Several days after the attacks, from the rise in *modularity* and the decrease in *Gini coefficient*, we can perceive the gradual recover process that the highly skewed allocation of collective attention recovers to the former versatile state but the residue of the shocking attacks still preserved in public's attention.

In general, the attention shift network could precisely reveal the attention shift process, exhibiting evident statistical and structural changes under exogenous shocks. Also, the network itself offers an unprecedented way to visualize the attention transferring process and some of the nested metrics show great potential of detecting real-time events which is a possible direction for further study.

V. Conclusion

In this paper, we present a novel approach to represent the collective attention shifting dynamics by extracting hashtags from tweet streams. We systematically analyze the "scattered–concentrated–dispersed" process of collective attention around a violent terrorist attack. Our entire framework gives a summary of the changing patterns of collective attention. We also give a list of metrics for systematically understanding the temporal changes. Some of the metrics exhibit great potential for event detection or giving the quantitative define of sever levels of disasters, which are opportunities for future works.

Acknowledgment

This work is part of the research supported from NSF grant #1423697, #1634944 and the CRDF at the University of Pittsburgh. Any opinions, findings, and conclusions or recommendations expressed in this material do not necessarily reflect the views of the funding sources. Xingsheng He also thanks the support of China Scholarship Council.

References

[1] L. Palen and K. M. Anderson, "Crisis informaticsnew data for extraordinary times," *Science*, vol. 353, no. 6296, pp. 224–225, 2016.

[2] A. Acar and Y. Muraki, "Twitter for crisis communication: lessons learned from japan's tsunami disaster," *International Journal of Web Based Communities*, vol. 7, no. 3, pp. 392–402, 2011.

[3] P. Burnap, M. L. Williams, L. Sloan, O. Rana, W. Housley, A. Edwards, V. Knight, R. Procter, and A. Voss, "Tweeting the terror: modelling the social media reaction to the woolwich terrorist attack," *Social Network Analysis and Mining*, vol. 4, no. 1, pp. 1–14, 2014.

[4] F. Vis, "Twitter as a reporting tool for breaking news: Journalists tweeting the 2011 uk riots," *Digital Journalism*, vol. 1, no. 1, pp. 27–47, 2013.

[5] K. Sasahara, Y. Hirata, M. Toyoda, M. Kitsuregawa, and K. Aihara, "Quantifying collective attention from tweet stream," *PloS one*, vol. 8, no. 4, p. e61823, 2013.

[6] J. Lehmann, B. Gonçalves, J. J. Ramasco, and C. Cattuto, "Dynamical classes of collective attention in twitter," in *Proceedings of the 21st international conference on World Wide Web*. ACM, 2012, pp. 251–260.

[7] S. D. Roy, G. Lotan, and W. K. Zeng, "The attention automaton: Sensing collective user interests in social network communities," *IEEE transactions on network science and engineering*, vol. 2, no. 1, pp. 40–52, 2015.

[8] C.-J. Wang and L. Wu, "The scaling of attention networks," *Physica A: Statistical Mechanics and its Applications*, vol. 448, pp. 196–204, 2016.

[9] F. Wu and B. A. Huberman, "Novelty and collective attention," *Proceedings of the National Academy of Sciences*, vol. 104, no. 45, pp. 17 599–17 601, 2007.

[10] K. Lee, J. Caverlee, K. Y. Kamath, and Z. Cheng, "Detecting collective attention spam," in *Proceedings of the 2nd Joint WICOW/AIRWeb Workshop on Web Quality*. ACM, 2012, pp. 48–55.

[11] Y.-R. Lin, B. Keegan, D. Margolin, and D. Lazer, "Rising tides or rising stars?: Dynamics of shared attention on twitter during media events," *PloS one*, vol. 9, no. 5, p. e94093, 2014.

[12] L. Weng, A. Flammini, A. Vespignani, and F. Menczer, "Competition among memes in a world with limited attention," *Scientific reports*, vol. 2, 2012.

[13] J. Yin, A. Lampert, M. Cameron, B. Robinson, and R. Power, "Using social media to enhance emergency situation awareness," *IEEE Intelligent Systems*, vol. 27, no. 6, pp. 52–59, 2012.

[14] Y. Sano, K. Yamada, H. Watanabe, H. Takayasu, and M. Takayasu, "Empirical analysis of collective human behavior for extraordinary events in the blogosphere," *Physical Review E*, vol. 87, no. 1, p. 012805, 2013.

[15] D. S. Bassett, D. L. Alderson, and J. M. Carlson, "Collective decision dynamics in the presence of external drivers," *Physical Review E*, vol. 86, no. 3, p. 036105, 2012.

[16] W. X. and L. Y.-R., "Sensing distress following a terrorist event," in *In 2016 International Conference on Social Computing, Behavioral-Cultural Modeling & Prediction and Behavior Representation in Modeling and Simulation (SBP-BRiMS 2016)*, 2016.

[17] M. Rosvall and C. T. Bergstrom, "Maps of random walks on complex networks reveal community structure," *Proceedings of the National Academy of Sciences*, vol. 105, no. 4, pp. 1118–1123, 2008.

[18] G. Fagiolo, "Clustering in complex directed networks," *Physical Review E*, vol. 76, no. 2, p. 026107, 2007.

[19] Y.-R. Lin, B. Margolin, A. Keegan, A. Baronchelli, and D. Lazer, "# bigbirds never die: Understanding social dynamics of emergent hashtag," in *Proceedings of the 7th International AAAI Conference on Weblogs and Social Media (ICWSM 2013)*, 2013.