How has Twitter changed the Event Discussion Scenario?: A Spatio-Temporal Diffusion Analysis

Purva Pruthi, Anu Yadav, Farheen Abbasi and Durga Toshniwal
Indian Institute of Technology Roorkee, Uttarakhand, India.

{purvapruthi617, anuyadav398, farheenabbasi47, durgatoshniwal}@gmail.com

***Abstract*—Earlier during the times of traditional print media, there used to be one-way information dissemination which was restricted to geographical boundaries having limited span and reach. With the advent of online social media, the process of information diffusion has changed significantly. It has become the fastest means of communication gaining wide popularity. Online Social Networks like Facebook, Twitter have revolutionized the interpersonal communication by providing a platform to individuals to express themselves at a global level, beyond their immediate geography. Most research in this area has focused on analyzing general information diffusion phenomenon. Our aim is to study diffusion dynamics of specific real world events, discussed on Twitter, with respect to location and time.We categorize the events into broad categories based on the following features - temporal (short or long), geo-spatial distribution (local or global), information diffusion mechanism (viral or gradual), influence(popular or unpopular) and cause (natural or planned). Temporal analysis shows that pre-event, during-event and post-event frequency distribution of tweets differ with respect to nature of events. For example, a planned event like "Delhi Elections" is more discussed after its actual occurrence whereas other planned event like "Obama's visit to India" is mainly discussed during the visit only. Through geospatial analysis, we find that some events which are supposed to be constrained locally, cross regional boundaries and become a matter of global discussion.We also study the diffusion of the events using the user interaction graph formed by retweet/mention links. We conclude with the three-dimensional analysis of spatio-temporal diffusion dynamics of real-world events by exploring relationships among them.**

***Index Terms*—Information Diffusion Analysis, Geo-spatial, Temporal, Real-world events, Online social media**

I. Introduction

Every moment, a lot of real-world events are happening all around the world. Earlier, during the times of printed social media it used to be very difficult to analyse the propagation of these events. With the advent of online social media and networking sites, it has become quite convenient to answer the questions like: How are these events spreading worldwide? What is the scope of influence - local or global? Are the events spreading virally or gradually?. Online social media provides a platform to the worldwide users, to be expressive and responsive about happenings instantly which has led to the generation of large amount of information giving rise to research fields like opinion mining in social media, sentiment detection or information diffusion analysis in online social networks.

There are differences in the mechanisms of information propagation across various kinds of events due to difference in underlying cause, concern, target population and inherent nature. For example events like floods or earthquake (natural disasters) may have a global, local or state-level concern depending on the scale of destruction. On the other hand, discussion about events like elections is expected to get trapped inside local or national community boundaries depending on the population affected due to results. Sometimes,they may also have a wider spread for discussion in case of controversies. Thus, actual diffusion of events in real-world differs from event-driven information propagation in online networks and there are various factors which determine the diffusion dynamics of real-world events in actual world as well as the virtual world.

Based on these premises, we investigate the following questions:
Q 1: Are local events discussed locally or globally?
Q 2: Do people discuss more during the events as compared to pre-event and post-event talks?
Q 3: Is India occupying an important position in the world?

The main contribution of this paper can be summarised in folllowing points covering 3 dimensional analysis:

- Spatial: Geospatial tweets distribution of different kinds of events is studied and we find that local and global patterns of events are also visible in the event discussion phenomenon on twitter.
- Temporal : Temporal analysis shows that pre-event, during-event and post-event frequency distribution of tweets differ with respect to nature of events. Long term and short term nature of events are described using temporal patterns.
- Information Diffusion : Finally, spatial-temporal analysis of retweet network for each event is performed and relationships between adoption time lag of retweets and their spatial distance are explored.

A. Organization of Paper:

The rest of the paper is organized as follows: Section 2 describes the related work done in the field of information diffusion analysis of events in Social Networks. In Section 3, we describe our Twitter dataset In Section 4, we discuss the experiments performed to explore the geospatial and temporal

978-1-4673-7278-7/15 $31.00 © 2015 IEEE
DOI 10.1109/BigDataCongress.2015.120

properties of the events and finally establish a relationship between sptial and temporal characteristics of various events and their observed diffusion patterns. Results are discussed in this section with possible explanations. Finally, we discuss the applications of our work in Section 6 followed by conclusion in the last section.

II. Related Work

Many measurement studies have been done to study the structural properties of social networks like Youtube,Myspace,Facebook,Twitter etc. [1,2,3,4] implying that different social networks exhibit different structural and behavioral properties. Our study focuses on microblogging social networking service, Twitter and is mainly based on the following lines of research:

Geo-spatial and Temporal analysis Geo-spatial analysis on twitter has enabled us to predict geographic location of users from tweets [5]. A similar analysis has been done in [6] which infers users location on Facebook based on their social network. Ardon et al. [7] studies how topics become popular in Twitter. They deduced that popular topics have one giant component that eventually covers a significant portion of network.Further ,geospatial analysis showed that highly popular topics cross regional boundaries. Researchers in [8] have used geo-tagged tweets to examine the impact of time, distance and location on the adoption of hashtags.They have also studied variation of hashtags with space by studying their properties like spread, entropy and focus.

The study of thematic based information diffusion processes with geo-spatial and temporal properties of data has remained unexplored till today. We aim at combining these aspects together. The influence of time and geographical location on information diffusion of specific events is our main concern. Generalisation of diffusion dynamics of these events may help in understanding of similar future events.

III. Dataset Description

Twitter provides a streaming API to extract tweets in real-time.We have collected public tweets related to various events by filtering tweets on specific keywords in twitter crawling API using tweepy library. The detailed description of the collected data is provided in Table 1.

Approximately 0.01 % of daily tweets on Twitter are geotagged containing latitude/longitude information. A socio-temporal study of geotagged tweets is done in [9] using 2 billion geotagged tweets containing 27 million unique hashtags collected over a span of 18 months. The events which we are dealing with do not last for such a long period of time.Thus, due to availabilty of very few fraction of geo-enabled tweets, significant spatial analysis was difficult for specific real-world events using only geotagged tweets. To obtain more spatial-information from the collected data, we used location information provided by the users on their twitter profile. Although user-specified location may not be that accurate and reliable as compared to the infromation provided by geo-enabled tweets, but performing such analysis gives an insight on the information-sharing behavior of users that how comfortable and serious they are in sharing their personal information. Also some users add invalidate locations like 'Anywhere on the Earth'. In order to use only valid locations, we geocoded the extracted locatinons using Data Science toolkit geocoding API. About 65 % of user-provided locations could be geocoded and used in analysis. Figure 1 shows the spatial distribution of geotagged tweets and geocoded tweets about Delhi election events. We observe that geocoded spatial distribution densify the regions already indicated by geoenabled tweets other than highlighting new areas.

TABLE I: Dataset Description

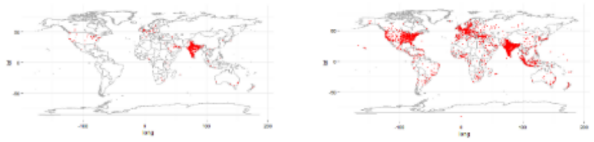
Event Name	Event Description	Hashtags	Number of Tweets	Duration of tweets extraction
Ebola disease	Outbreak of ebola virus epidemic disease in parts of South Africa in 2014	#ebola	0.69 million	November 5th 2014 to November 22nd 2014
Terror attack in Paris	12 Cartoonists of French satirical MagazineCharlie Hebdo were killed by group of Terrorists on 7th Jan 2014. People strongly criticised these attacks on social media	#CharlieHebdo, #JesuisCharlie	0.41 million	January 13,2015-Feb 3rd, 2015
Barack Obama's visit in India	US President had a 3-day visit in India on the occasion of India's Republic Day,26th Jan 2015	#ObamaInIndia, #NamasteObama	0.24 million	January 13,2015 to Feb 3rd, 2015
Delhi Elections	State elections in Delhi,India were held on 7th Feb,2015. It was a very crucial event with people actively expressing their poll opinions on social media	#DelhiElections, #DelhiNext, #DelhiPolls	0.2 million	Jan 14th 2015 to Feb 15th 2015
India-Pakistan Cricket World-cup Match	Cricket Match held between India and Pakistan in Adelaide,Australia on 15th Feb,2015	#INDvsPAK	0.76 million	Feb 14th 2015 to Feb 19th 2015
Fifty shades of grey movie release	The movie released on Feb 14th,2015 and faced many controversies all around the world.	#50ShadesOfGrey	0.29 million	Feb 6th,2015 to Feb 17th, 2015

IV. Experiments And Results

A. Geospatial Properties of Events

We define following terminology with respect to the locations associated with events.

Origin Location: It is the actual location where the event took place. For example in case of Obama's visit in India, origin location is Delhi, India where he actually visited and paticipated in Republic Day Celebration.



(a) Geotagged Tweets (b) Geocoded Tweets

Fig. 1: Geospatial Distribution of tweets on Delhi Elections in India

Concerned Location : The location whose population is concerned with the impact of event's occurence. Note that, concerned locations may be other than the origin location.

In this section, we begin our analysis by examining the spatial distribution of events across countries and exploring the relationship between the origin location and concerned location of events and the locations where they are discussed. In particular, we are interested in understanding: (i) what is the worldwide distribution of different events? (ii) How does the cumulative frequency distribution of tweets vary as we move away from origin locations (iii) If an event is associated with multiple locations which kind of location dominates in the discussion of event - concerned locations or origin locations. Initially, we perform an exploratory analysis by mapping lat/long of tweets on map and visualizing the spread. In order to quantify the distribution of tweets with respect to the origin of event, we calculate the great circle distance between origin location or other concerned locations and tweet's user location using Haversine Method,which accounts for the effects of the Earths spherical shape in finding distances between points.We call this distance difference *distdiff* and plot it against the cumulative frequency of tweets observed within that distance. In Figure 2, we observe that discussion about India-Pakistan Match is most dominant in India as we observe steep slope of graph around distance 0 km if *distdiff* is calculated taking India as reference location. Similarly, we observe that in Distance from Adelaide graph, discussion is low around regions of Australia but it starts rising at a distance of 8000km which is approximately same as distance between Adelaide and India (8,806 km). As India and Pakistan are neighbouring countries, we observe that *Distdiff* from Pakistan as a reference location is similar with India graph but slope is less depicting that discussion is more in Indian regions as compared to Pakistan regions.

Figure 3 shows the distribution of two similar kind of events- Political Leaders' visit to foreign countries. In Figure 3(a), India is host country for US President Barack Obama. It shows tha Indian people are more excited about Obama's visit as compared to USA population. Similar analysis on Indian Prime Minister's visit in Australia shows the reverse situation where hosting population is not talking much about visit on Twitter as compared to Indian people. It shows the active behavior of Indian population,in general, on twitter regarding Poplitical leader's visit.

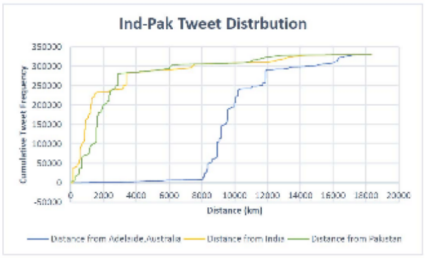
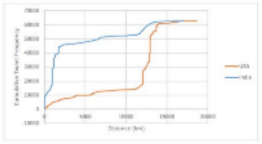
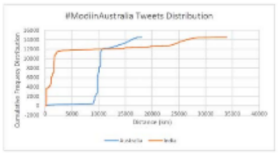


Fig. 2: Ind-Pak Tweets Distribution from Various Associated Locations



(a) #ObamainIndia Tweets Distribution from USA and India



(b) #ModiinAustralia Tweets Distribution from Australia and India

Fig. 3: Tweets Distribution of two Political Leaders' visit from visiting country and leader's original country

The events which we have discussed till now share a common characteristic that they are more of interest to the regional population. These events are inherently local in nature and are not drawing the global attention. Figure 4 shows the *distdiff* graphs for globally concerned events like Ebola spread, Charlie Hebdo Terror attack and Release of Movie-Fifty shades of Grey, taking their respective origin locations as reference. Frequency peaks are not observed in these graphs and shows the smooth distribution of tweets across the world. The similar shapes of curves for these events confirms the global importance of these events.

Therefore, in this section we observed the differences in the spatial characteristics of a variety of events on Twitter. We conclude that the events which seem of local influence are actually spreading in similar fashion on Twitter with high peaks in concerned local regions whereas the events which originally seeem of global importance have smooth global spread across all the regions of the world.

B. Temporal Properties of Events

In this section, we analyse the temporal distribution of events over the time duration when tweets were collected. Figure 5 shows the temporal variation of tweets, red vertical lines highlight the actual time duration when event occured. Pre-event discussion refers to the left of vertical lines and post-event discussion refers to the right of red vertical lines. It shows that pre-event, during-event and post-event frequency distribution of tweets differ with respect to nature of events and spikes in frequency are observed during controversial moments during the events.

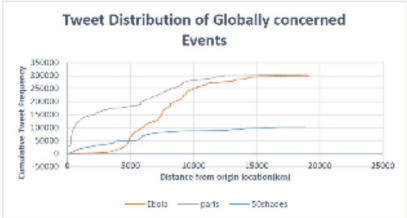
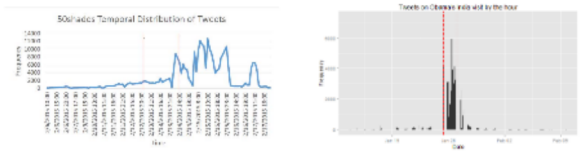


Fig. 4: Spatial Distribution of Globally concerned events with respect to origin location



(a) 50 shades of Grey tweets frquency per min

(b) #ObamainIndia tweets frequency per hour

Fig. 5: Temporal Distribution

C. Spatio-Temporal Diffusion Analysis of Events using Retweet Interaction Networks Between Users

The time difference between a retweet and its original tweet is defined as adoption time lag of a retweet. In figure 6, adoption time difference of a retweet is plotted against the geographic distance between retweeting user and original user who posted the tweet. It is observed that most of the tweets are retweeted within 35 days. Almost similar graphs were obtained for retweet networks of other events.

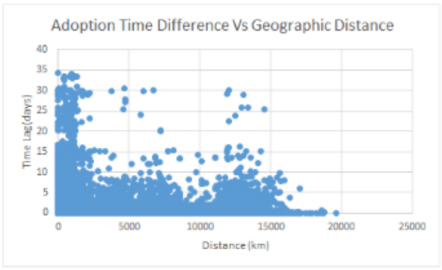


Fig. 6: Time difference in tweet adoption between users in Delhi's Election retweet network.

V. Applications

This kind of study helps in improving marketing campaigns by identifying the target customers. It helps in prediction of diffusion dynamics of similar kinds of future events. It plays a major role in analyzing how information propagates in social media sites. Geospatial diffusion study helps in the prediction of locations where event is most likely to be discussed in near future. Temporal diffusion analysis helps in understanding if an event is spreading gradually or virally and is used in viral event prediction. It is also used in understanding the evolution of event-related diffusion networks over time and finding time-series patterns. Thus this kind of study helps in the diffusion dynamics analysis of similar kinds of events and predict their outreach and spread.

VI. Conclusion

The research shows that spatial and temporal diffusion of events depend entirely on their nature. It is obsevered that supposedly local events like Delhi Elections, which seemed to be confined to ia only, are discussed on a much larger scale. It is also found that discussion of certain events was not limited to origin of events. For example, first outbreak of Ebola disease took place in South Africa, but it was more discussed in USA and other countries. Similarly one day event like INDvsPAK match despite being played in Australia is more discussed in India and Pakistan. Also, temporal diffusion patterns of events vary according to characteristics of events like if it is planned or sudden. Therefore, spatio-temporal analysis of distribution of event related tweets on Twitter gives us an insight into the characteristics exhibited by different event discussion phenomenon in social media networks. As part of future work, we intend to analyze the temporal and spatial evolution of diffusion graphs using temporal metrics for time-varying graphs.

References

[1] Mislove, Alan, et al. "Measurement and analysis of online social networks." Proceedings of the 7th ACM SIGCOMM conference on Internet measurement. ACM, 2007.

[2] Ahn, Yong-Yeol, et al. "Analysis of topological characteristics of huge online social networking services." Proceedings of the 16th international conference on World Wide Web. ACM, 2007.

[3] Wilson, Christo, et al. "User interactions in social networks and their implications." Proceedings of the 4th ACM European conference on Computer systems. Acm, 2009.

[4] H. Kwak, C. Lee, H. Park, and S. Moon. What is twitter, a social network or a news media? WWW10.

[5] Z. Cheng, J. Caverlee, and K. Lee. You are where you tweet: a content-based approach to geo-locating twitter users.CIKM 10

[6] [L. Backstrom, E. Sun, and C. Marlow. Find me if you can: improving geographical prediction with social and spatial proximity. WWW 10.

[7] Sebastien Ardon, Amitabha Bagchi, Anirban Mahanti, Amit Ruhela, Aaditeshwar Seth, Rudra M. Tripathy, and Sipat Triukose. Spatio-temporal analysis of topic popularity in twitter. CoRR, abs/1111.2904, 2011.

[8] D. Romero, B. Meeder, and J. Kleinberg. Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter. In Proceedings of the 20th international conference on World wide web, pages 695704. ACM, 2011.

[9] Kamath, Krishna Y., et al. "Spatio-temporal dynamics of online memes: a study of geo-tagged tweets." Proceedings of the 22nd international conference on World Wide Web. International World Wide Web Conferences Steering Committee, 2013.