

Traces of Social Media Activism from Malaysia and Pakistan

M. Atif Qureshi^{*†◊}, Arjumand Younus^{*†◊}, Lay-Ki Soon[‡], Muhammad Saeed[†], Nasir Touheed[†], Colm O’Riordan^{*}, Pasi Gabriella[◊]

^{*}Computational Intelligence Research Group, Information Technology, National University of Ireland, Galway, Ireland

[‡]Faculty of Information Technology, Multimedia University, Cyberjaya, Malaysia

[◊]Information Retrieval Lab, Informatics, Systems and Communication, University of Milan Bicocca, Milan, Italy

[†]Web Science Research Group, Faculty of Computer Science, Institute of Business Administration, Karachi, Pakistan

maqureshi, ayounus@iba.edu.pk, lksoon@mmu.edu.my, saeed@uok.edu.pk, ntouheed@iba.edu.pk, colm.oriordan@nuigalway.ie, pasi@disco.unimib.it

ABSTRACT

Recent uprisings in North Africa and Middle East saw the rise of social media as an online activism medium used frequently by activists. Social media activism is an evolving phenomenon, and it is viewed by many as the main gathering point in recent protests (e.g., Arab Spring, Occupy Wall Street). Recently, the research community has begun to investigate the intricate interplay between activism and social media. This paper presents a preliminary study in this direction by making use of publicly available Twitter data. We analyse the regional characteristics of online activists of two regions, Malaysia and Pakistan. Unlike previous approaches, we perform a content analysis of activists’ tweeting habits. For this, we identify the popular topics in both regions along with the frequency with which these topics are discussed by the activists according to their tweeting habits. Furthermore, we analyse the diversity of the content being shared by an activist with respect to their tweeting habits from both regions. Our results show differences in tweeting habits across the two regions for a period of 45 days (Dec. 1st, 2011 to Jan. 15th, 2012).

Categories and Subject Descriptors

J.4 [Social And Behavioral Sciences]: Sociology; H.1.2 [User/Machine Systems]: [Human factors]

Keywords

activism, Twitter, popular topics, user behaviour

1. INTRODUCTION

Social media applications have considerably influenced the traditional paradigm of journalism giving birth to the phenomenon of social media activism. The term social media activism includes a broad range of activities from real-time news publishing to organization of large-scale protest. The highlighting characteristic is that all these activities are performed through social media platforms. The recent uprisings in the Arab world witnessed active use of social media activism and since then there has been an increasing use of social media platforms for activism.

Additionally, the paradigm shift in journalism has attracted much attention in the research community. There have been many studies to investigate the inner workings of social media and various phenomenon behind it. Our work differs from previous studies in that it provides a regional characterization of social media activism in one of the most popular social media namely Twitter. Despite the widespread adoption of Twitter as a social media activism platform, little is known about cross-regional differences in Twitter. To the best of our knowledge, our work is the first to systematically study how social media activists of different regions behave in Twitter. We address two main questions:

1. What behavioural differences are exhibited by social media activists of different regions?
2. Are there any commonalities between the topics shared by social media activists of different regions? What is the pattern in which social media activists of various regions share the top topics?

To address these questions, we performed an analysis of well-known Twitter activists from Malaysia and Pakistan. These activists were manually sampled by the authors from those respective countries: a total of 70 Malaysian Twitter activists and 118 Pakistani Twitter activists were collected which resulted in the collection of a total of 183,028 tweets. First, we identify the topics discussed by Twitter activists of these two regions. Second, we show significant differences

across these topics and in the tweeting activities of activists from these two regions.

The remainder of this paper is organized as follows. Section 2 presents a brief overview of related work. Section 3 lays out the basic features of Twitter. Section 4 describes our experimental methodology along with details of the characteristic differences between the social media activists from the two regions. Section 5 concludes the paper with a discussion on implications of this research.

2. RELATED WORK

Many studies have provided insights into user participation in social media along various dimensions. In this section, we provide a brief overview of some previous studies and also highlight how our work differs from these studies.

Early research within the domain characterized the ways in which people used social media platforms [5, 9]. Later, researchers started to investigate various other phenomena within social media such as information diffusion and news propagation [10, 6], influence analysis [2, 4], and news detection [7, 8]. In short, most of the published research on social media has focused on questions of a general nature without taking into account its nature of being an online media outlet. However, a recent work by An et al.[1] pursues a study of Twitter as a media platform and presents a study of the media landscape in Twitter. Their methodology is built upon network-based techniques and much of their findings come through analysis of social links in Twitter. Less work has presented a systematic analysis of the textual content of Twitter posts. This paper attempts to fill that gap and aims to characterize the journalistic behaviour in terms of the content being shared on social media.

3. TWITTER CHARACTERISTICS

Twitter is a unique platform in that it has been termed as both a micro-blogging medium as well as a social network. Its micro-blogging nature comes from its status update feature which enables users able to answer the question of ‘What’s happening?’ in 140 characters. In Twitter, social links are created through the follower/following paradigm where Twitter users can follow others or can be followed. Unlike most online social networking sites this concept does not require any reciprocation. A user can follow any other user, and the user followed does not necessarily have to follow back. Being a follower on Twitter implies that the user receives all the messages (called tweets) from those the user follows. A well-defined practice of conversation on Twitter is through a mention: ‘@’ followed by a user identifier addressing the mentioned user. Another feature namely RT empowers users to spread information received by one’s followers: RT stands for a retweet and it refers to the practice of sharing the same content as that shared by another Twitterer - this is also another mode of replying on Twitter. Another special feature of tweets is that they sometimes contain the hash symbol ‘#’ referred to as hashtags. By including a hashtag in a tweet, the user originating the tweet is suggesting that the word denoted by the hashtag makes for a good candidate as a search key for the tweet. We refer to these type of tweets that Twitter allows (namely, statuses, mentions, and retweets) as posting types and present a characterization of frequency use of each by Twitter activists of

Country	Malaysia	Pakistan
Mean no. of tweets per activist	204	621
Mean no. of followers per activist	2149	2405
Maximum no. of followers per activist	44642	80890
Minimum no. of followers	53	208

Table 1: Basic Twitter statistics for Malaysian and Pakistani social media activists

both Malaysia and Pakistan in the next section.

4. JOURNALISTS’ BEHAVIOUR ON TWITTER

In this section we describe our approach for user profiling journalists from Malaysia and Pakistan. We also present results of our analysis. As mentioned in Section 1, we collected tweets of Pakistani and Malaysian social media activists. We collected the tweet data for a period of one and a half months starting December 01, 2011 to January 15, 2011. Table 1 shows some basic statistics about the Twitter activists of both the countries. As shown in Table 1, there is a wide disparity in the tweeting activity of journalists from the two countries with Pakistani Twitter activists being highly active.

Figure 1 and 2 show the top posting types of Malaysian and Pakistani activists respectively. As seen from the Figures, Malaysian and Pakistani activists exhibit a significant difference in their mentions activity with Pakistanis using this feature of Twitter very frequently implying that Pakistani activists mostly use Twitter for conversations with each other. Secondly, Pakistani Twitter activist endorse each other in the form of retweets at a higher frequency than that of Malaysians. Another difference is in the status updates and it clearly shows that Pakistani activists share their thoughts less frequently through status updates while Malaysians update the status on a frequent basis. Some similarities exist between the two groups in that mentions are by far the most common posting type and in that retweets are the least frequent. These figures demonstrate that Pakistani Twitter activists engage in conversations (i.e., mentions and retweets) very frequently.

4.1 Experimental Methodology

The approach we use for topic identification in tweet content is based on a latent variable topic model, namely Latent Dirichlet Allocation [3]. It is an unsupervised machine learning technique which uncovers information about latent topics across a corpora. We use a special variant of LDA (namely Twitter-LDA) proposed by Zhao et al.[11]. This model has been adopted for Twitter and is based on the assumption that a single tweet is about a single topic (this assumption makes sense given the 140-character limit in Twitter).

Application of Twitter-LDA on the tweets’ dataset yields the topics of interest of Twitter activists and the words associated with these topics. We then perform tweet cluster-

¹Note that we translate the Malaysian tweets into English using our own translation system.

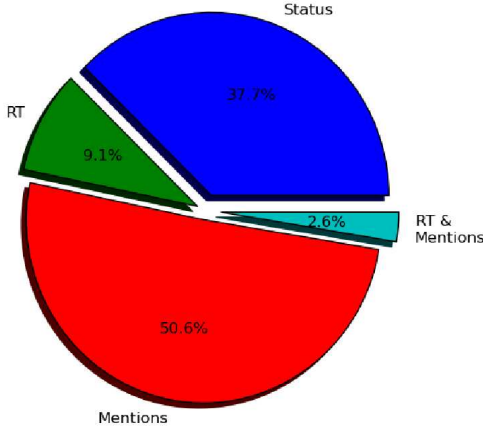


Figure 1: Top posting types of Malaysian activists

ing with respect to discovered topics along the dimensions of statuses, mentions, retweets, mention endorsements and retweet endorsements. Statuses, mentions and retweets are standard ways to share content on Twitter. We define mention endorsements and retweet endorsements for an activist as follows:

- Mention endorsements of an activist refers to when the activist is mentioned by other activists
- Retweet endorsements of an activist refers to when the activist's tweet is retweeted by other activists

We perform two types of studies on the tweets of the activists. The first one which we refer to as a popular topics study investigates the top topics in statuses, mentions, retweets, mention endorsements and retweet endorsements. The second one which we refer to as a user behaviour study investigates the diversity of topics in activists' statuses, mentions, retweets, mention endorsements and retweet endorsements. Both these' investigations are performed as follows.

Every set of tweets involving an activist (either produced by him/her or endorsed by others to him/her) is first decomposed into a matrix of statuses (shortly, s), mentions (shortly, m), retweets (shortly, rt), mention endorsements (shortly, m_e) and retweet endorsements (shortly, rt_e) along 'n' discovered topics. A typical example is as follows.

$$\begin{bmatrix} s_1 & s_2 & \dots & s_i & \dots & s_n \\ m_1 & m_2 & \dots & m_i & \dots & m_n \\ rt_1 & rt_2 & \dots & rt_i & \dots & rt_n \\ m_e_1 & m_e_2 & \dots & m_e_i & \dots & m_e_n \\ rt_e_1 & rt_e_2 & \dots & rt_e_i & \dots & rt_e_n \end{bmatrix}$$

Here s_i denotes the number of statuses referring to a topic 'i'

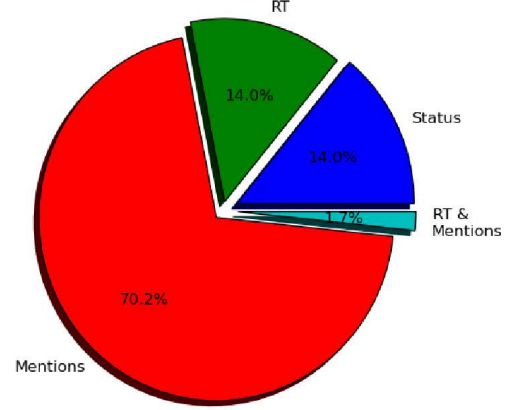


Figure 2: Top posting types of Pakistani activists

by a particular activist, m_i denotes the number of mentions referring to a topic 'i' by a particular activist and so on.

The generic notation we use is that of a topic vector T shown as follows:

$$[T_1 \ T_2 \ \dots \ T_i \ \dots \ T_n]$$

For the popular topics study, we first determine the set of top 'm' topics for each activist's statuses (s_{top_m}), mentions (m_{top_m}), retweets (rt_{top_m}), mention endorsements ($m_e_{top_m}$) and retweet endorsements ($rt_e_{top_m}$). Finally, for each topic t_i in T, we determine its popularity, $P[i]$, by counting the number of intersections among all activists' top 'm' topics for each of statuses, mentions, retweets, mention endorsements and retweet endorsements as follows:

$$\sum_{\forall \text{activists}} \begin{cases} \text{add } 1 & \text{if } t_i \in \mathbb{T}_{u_{top_m}} \\ 0 & \text{otherwise} \end{cases}$$

Similarly for the user behaviour study, first we determine the set of top 'm' topics for each activist's statuses (s_{top_m}), mentions (m_{top_m}), retweets (rt_{top_m}), mention endorsements ($m_e_{top_m}$) and retweet endorsements ($rt_e_{top_m}$). Then we took intersections between top 'm' topics of a combination of tweet types for each activist. The combination of tweet types we use are statuses and retweets (shortly, $s-r$), statuses, mentions and retweets (shortly, $s-m-rt$), statuses and mentions (shortly, $s-m$), retweets and mentions (shortly, $rt-m$), mention endorsements and retweet endorsements (shortly, m_e-rt_e), retweets and retweet endorsements (shortly, $rt-rt_e$), mentions and mention endorsements (shortly, $m-m_e$), statuses and retweet endorsements (shortly, $rt-rt_e$), and finally, statuses and mention endorsements (shortly, $rt-rt_e$).

Top topics in tweets of Malaysian activists		
Popular Topic Among Top 'm' Tweet Type	Tweet Type & Frequency	Topic
m=1	Statuses (24) Mentions (31) Retweets (19) Mention Endorsements (15) Retweet Endorsements (8)	39, quot, umno, sonny, kl, malay, people, izzah 39, quot, good, hehe, christmas, posted, happy, bro 39, quot, umno, sonny, kl, malay, people, izzah 39, quot, good, hehe, christmas, posted, happy, bro 39, quot, umno, sonny, kl, malay, people, izzah
m=2	Statuses (39) Mentions (40) Retweets (35) Mention Endorsements (20) Retweet Endorsements (8)	39, quot, umno, sonny, kl, malay, people, izzah 39, quot, good, hehe, christmas, posted, happy, bro 39, quot, umno, sonny, kl, malay, people, izzah 39, quot, good, hehe, christmas, posted, happy, bro 39, quot, cpb, mengsk, movie, ahfugs, sagittarius
m=3	Statuses (45) Mentions (50) Retweets (38) Mention Endorsements (20) Retweet Endorsements (8)	39, quot, umno, sonny, kl, malay, people, izzah 39, quot, good, hehe, christmas, posted, happy, bro 39, quot, umno, sonny, kl, malay, people, izzah 39, quot, good, hehe, christmas, posted, happy, bro 39, quot, cpb, mengsk, movie, ahfugs, sagittarius
m=4	Statuses (48) Mentions (53) Retweets (40) Mention Endorsements (20) Retweet Endorsements (8)	39, quot, umno, sonny, kl, malay, people, izzah 39, quot, good, hehe, christmas, posted, happy, bro 39, quot, umno, sonny, kl, malay, people, izzah 39, quot, good, hehe, christmas, posted, happy, bro 39, quot, cpb, mengsk, movie, ahfugs, sagittarius

Table 2: Tweeting habits of Malaysian activists

Top topics in tweets of Pakistani activists		
Popular Topic Among Top 'm' Tweet Type	Tweet Type & Frequency	Topic
m=1	Statuses (12) Mentions (12) Retweets (14) Mention Endorsements (25) Retweet Endorsements (12)	day, time, work, good, karachi pakistan, thingsdesiparentssay, sir, par, aaj pakistan, taliban, killed, karachi, drone good, ff, dont, twitter, yeah, time pakistan, ara, blog, pba, media
m=2	Statuses (18) Mentions (24) Retweets (25) Mention Endorsements (47) Retweet Endorsements (18)	day, time, work, good, karachi good, ff, dont, twitter, yeah, time pakistan, taliban, killed, karachi, drone good, ff, dont, twitter, yeah, time pakistan, isi, army, coup, mansoor
m=3	Statuses (25) Mentions (41) Retweets (35) Mention Endorsements (66) Retweet Endorsements (22)	day, time, work, good, karachi good, ff, dont, twitter, yeah, time pakistan, taliban, killed, karachi, drone good, ff, dont, twitter, yeah, time pakistan, isi, army, coup, mansoor
m=4	Statuses (30) Mentions (51) Retweets (46) Mention Endorsements (70) Retweet Endorsements (27)	day, time, work, good, karachi good, ff, dont, twitter, yeah, time pakistan, taliban, killed, karachi, drone good, ff, dont, twitter, yeah, time pti, pakistan, imran, khan, ik

Table 3: Tweeting habits of Pakistani activists

Finally, we determine user behaviour by counting the number of activists that have an intersection in top 'm' topics for each tweet type combination as follows:

$$\sum_{\forall \text{activists}} \begin{cases} \text{add } 1 & \text{if } |Tu_{top_m} \cap Tu_{top_m}| > 0 \\ 0 & \text{otherwise} \end{cases}$$

Similar to the popular topics' study, this phase is also sepa-

rately performed for Twitter activists of both regions so as to identify activists' diversity for both Malaysia and Pakistan. Results of this analysis are shown in the next subsection.

4.2 Experimental Results

In this Section, we present the results of our analysis. Table 2 and 3 show the top topics for both Malaysia and Pakistan. The results are shown by varying m of the previous section up to a value of four (e.g., $m=1$ implies the top one topic in the various tweet types, $m=2$ implies top two

topics in the various tweet types and so on). We show the results separately for statuses, mentions, retweets, mention endorsements and retweet endorsements. We show the top eight words of each topic for Malaysian Twitter activists and top five words of each topic for Pakistani Twitter activists: more words are chosen for Malaysian tweets due to the ambiguity issues that arise when less words are taken. From Table 2 and 3 it becomes evident that for statuses and mentions the nature of Pakistani Twitter activists' topics is personal (comprising of words such as day, time, work, good, karachi, ff, dont, twitter, yeah) whereas retweets along with retweet endorsements are of a political nature (comprising of words such as pakistan, taliban, killed, karachi, drone, pti, pakistan, imran, khan, isi, army, coup, mansoor). On the other hand, for statuses and retweets the nature of Malaysian Twitter activists' topics is political (comprising of words such as umno, kl, malay, people, izzah) whereas mentions reflect personal chatting activity (comprising of words such as good, hehe, christmas, posted, happy). For the duration of the data collection, many significant events took place in Pakistani politics due to which we see a large number of political words (i.e., pakistan, taliban, killed, karachi, drone, pti, pakistan, imran, khan, isi, army, coup, mansoor); on the other hand, significant events also took place in Malaysian politics but most of them were related to a umno (United Malays National Organization) and Nurul Izzah (daughter of Malaysia's prominent opposition leader, Anwar Ibrahim).

Figure 3 and 4 show the ratio of Malaysian and Pakistani activists that share common topics in the various tweet type combinations respectively. As clear from Figure 3, Malaysians exhibit greater diversity in their tweeting habits. According to our user behaviour analysis, 75.32% of Malaysian activists have at least one common topic among top four topics in statuses and retweets, 71.43% of Malaysian activists have at least one common topic among top four topics in statuses, retweets and mentions. On the other hand, 91.74% of Pakistani Twitter activists have at least one common topic among top four topics in statuses and retweets and 84.30% of Pakistani Twitter activists have at least one common topic among top four topics in statuses, retweets and mentions. The complete analyses can be seen from Figure 3 and Figure 4. The high ratios for Pakistani activists denotes the similarity of their topics across the various tweeting activities (i.e., statuses, mentions, retweets, mention endorsements and retweet endorsements).

5. CONCLUSION

In this paper, we showed differences in social media activism (more specifically, tweeting habits) of Pakistani activists and Malaysian activists. Such analysis is extremely valuable in that the proposed content analysis features can be used to study evolution of political campaigns in tweeting habits, such as how a particular topic is being projected or popularized. We aim to pursue such a study as future work. Similarly, it can also help in understanding personalized preferences of an activist with regards to his/her topics of interest. This can aid in improving personalized search and personalized recommendation systems of the future. We hope that this paper represents a step towards understanding how various cultures and nations perceive social media activism and how they interact with social media platforms in the future. We hope to extend our study to understand the dynamics of

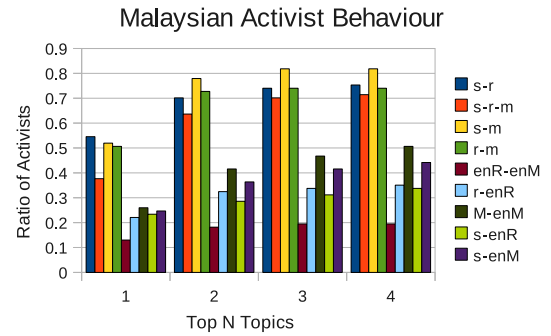


Figure 3: Malaysian activists' diversity: common topics across tweet type combinations for top 'n' topics

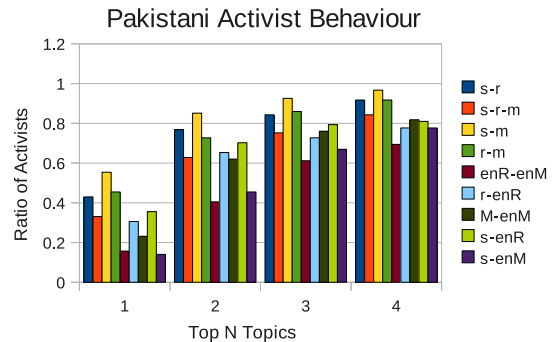


Figure 4: Pakistani activists' diversity: common topics across tweet type combinations for top 'n' topics

social media activism by repeating the study for a longer period of time in order to observe tweeting habits independent of few events (which could cause a bias in analysis).

6. REFERENCES

- [1] J. An, M. Cha, K. Gunmadi, and J. Crowcroft. Media landscape in Twitter: A world of new conventions and political diversity. In *Proceedings of 5th International AAAI Conference on Weblogs and Social Media*, pages 18–25, 2011.
- [2] E. Bakshy, J. M. Hofman, W. A. Mason, and D. J. Watts. Everyone's an influencer: quantifying influence on twitter. In *Proceedings of the fourth ACM international conference on Web search and data mining, WSDM '11*, pages 65–74, New York, NY, USA, 2011. ACM.
- [3] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3:993–1022, Mar. 2003.
- [4] M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gummadi. Measuring user influence in twitter: The million follower fallacy. In *in ICWSM '10: Proceedings of international AAAI Conference on Weblogs and Social*, 2010.

- [5] C. Honey and S. Herring. Beyond microblogging: Conversation and collaboration via twitter. In *System Sciences, 2009. HICSS '09. 42nd Hawaii International Conference on*, pages 1–10, jan. 2009.
- [6] D. M. 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, WWW '11*, pages 695–704, New York, NY, USA, 2011. ACM.
- [7] T. Sakaki, M. Okazaki, and Y. Matsuo. Earthquake shakes twitter users: real-time event detection by social sensors. In *Proceedings of the 19th international conference on World wide web, WWW '10*, pages 851–860, New York, NY, USA, 2010. ACM.
- [8] J. Sankaranarayanan, H. Samet, B. E. Teitler, M. D. Lieberman, and J. Sperling. Twitterstand: news in tweets. In *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, GIS '09*, pages 42–51, New York, NY, USA, 2009. ACM.
- [9] F. Schneider, A. Feldmann, B. Krishnamurthy, and W. Willinger. Understanding online social network usage from a network perspective. In *Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference, IMC '09*, pages 35–48, New York, NY, USA, 2009. ACM.
- [10] J. Yang and J. Leskovec. Modeling information diffusion in implicit networks. In *Data Mining (ICDM), 2010 IEEE 10th International Conference on*, pages 599–608, dec. 2010.
- [11] W. Zhao, J. Jiang, J. Weng, J. He, E.-P. Lim, H. Yan, and X. Li. Comparing twitter and traditional media using topic models. In P. Clough, C. Foley, C. Gurrin, G. Jones, W. Kraaij, H. Lee, and V. Mudoch, editors, *Advances in Information Retrieval*, volume 6611 of *Lecture Notes in Computer Science*, pages 338–349. Springer Berlin / Heidelberg, 2011.