2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)

Characterizing Communal Microblogs during Disaster Events

Koustav Rudra
IIT Kharagpur, India
koustav.rudra@cse.iitkgp.ernet.in

Ashish Sharma
IIT Kharagpur, India
ashishshrma22@gmail.com

Niloy Ganguly
IIT Kharagpur, India
niloy@cse.iitkgp.ernet.in

Saptarshi Ghosh
IIEST Shibpur, India
sghosh@cs.iiests.ac.in

***Abstract*—Millions of microblogs are posted during disasters, which include not only information about the present situation, but also the emotions / opinions of the masses. While most of the prior research has been on extracting situational information, this work focuses on a particular type of non-situational tweets – communal tweets, i.e., abusive posts targeting specific religious / racial groups. We characterize the communal tweets posted during five recent disaster events, and the users who posted such tweets. We find that communal tweets are posted not only by common users, but also by many popular users (having tens of thousands of followers), most of whom are related to the media and politics. As a result, communal tweets get much higher exposure (retweets) than non-communal tweets. Further, users posting communal tweets form strong connected groups in the social network. Considering the potentially adverse effects of communal tweets during disasters, we also indicate a way to counter such tweets, by utilizing anti-communal tweets posted by some users during such events.**

I. Introduction

One of the ominous fallouts of disaster is the general depletion of spirit among the affected population. Often, taking advantage of such a vulnerable situation, hatred and misinformation are spread in the affected zone, which may result in serious deterioration of law and order situation. The spread of hatred is significantly enhanced through social media where Twitter is increasingly used as a powerful tool. Especially harmful and potentially dangerous are the *communal tweets*, which are directed towards certain religious or racial communities.

It has been observed earlier that such tweets are often posted during man-made disasters, such as terrorist attacks. For instance, Burnap *et al* [1] observed that during the Woolwich attack, the UK masses targeted a certain religious community to which the attackers belonged. However, we surprisingly observe that communal tweets are also posted during *natural disasters* like floods and earthquakes, at least in certain geographical regions such as the Indian subcontinent.

In this work, we first identify communal tweets using the methodology proposed in [1], and then study the nature of communal tweets, and the users who post them (Section IV). We find that, alarmingly, communal tweets are posted not only by common, random users but also by some very popular users, and such tweets are retweeted more frequently than other types of tweets. Interestingly, most of these popular users who post communal tweets, belong to either media houses or are in politics. These communal users develop a strong social network among themselves separated from non-communal users.

To our knowledge, this study is the first on characterizing communal tweets and users who posted such tweets during disasters, and it gives a novel insight into how social media platforms are used to spread communal content even during natural disasters in some regions. We also indicate a potential way of countering the spread of such communal content. We observe that a small number of users post *anti-communal content* which aim to maintain peace and harmony. However, such anti-communal content usually receives lot less exposure than the communal content. Promoting the anti-communal content can be a promising way to counter the communal venom posted during disasters.

II. Related Work

Online forums are increasingly being used by the masses to post hate speeches and offensive content. Hence, there have been lot of effort in recent years for automatic identification of such offensive content [1], [2], [3]. For instance, Greevy *et al.* [4] classified racist content in webpages using a supervised bag-of-words model. Dinakar *et al.* [5] identified *cyberbullying*, using features like profane words, parts-of-speech tags, words with negative connotations, and so on. Similarly, Chen *et al.* [6] used profanities, obscenities, and pejorative terms as features with appropriate weightage to identify offensive content in Youtube comments. Mahmud *et al.* [7] identified insulting syntactic constructs, relationship between terms to detect online flaming behaviour. More recently, Burnap *et al.* [1], [8] attempted to detect hate speech posted during a disaster event (the Woolwich attack).

Almost all prior works have focused on identifying offensive content and hate speech, and there has been very few efforts towards characterizing the users who post such contents. To the best of our knowledge, the recent study by Silva *et al* [3] is the only one which attempted to identify the sources and targets of such hate speeches. However, there has not been any detailed effort in characterizing users who post such content, especially during disaster events. In this work, we take the first step in this direction by characterizing the users posting communal tweets based on their popularity, interests, and social interactions.

IEEE/ACM ASONAM 2016, August 18-21, 2016, San Francisco, CA, USA
978-1-5090-2846-7/16/$31.00

TABLE I: Statistics of data collected.

Event	# Tweets	# Distinct users
NEQuake	5,05,077	3,26,536
KFlood	14,922	8,367
GShoot	53,807	29,293
PAttack	6,48,800	5,77,888
CShoot	2,93,483	1,64,276

III. Dataset and Methodology

Disaster events: We considered tweets posted during the following recent disaster events – (i) **NEQuake:** a devastating earthquake in Nepal, (ii) **KFlood:** floods in the state of Kashmir in India, (iii) **GShoot:** a terrorist attack in Gurudaspur, India, and (iv) **PAttack:** coordinated terrorist attacks in Paris, and (v) **CShoot:** terrorist attack at the Inland Regional Center in San Bernardino, California. Note that the first two events are natural disasters, while the latter three are man-made disasters. Additionally, we have selected events occurring in different geographical regions so that the study would not be biased to any particular demographics.

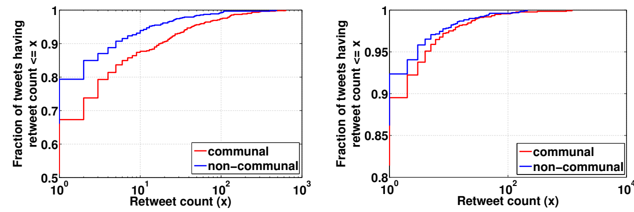
Collecting tweets posted during the disaster events: We collected relevant tweets posted during each event through the Twitter Search API [9] using keyword matching. For example, the keywords '#NepalEarthquake', 'Nepal' and 'quake' were used to identify tweets related to the NEQuake event, and for each keyword, all tweets returned by the Twitter Search API were collected. Subsequently, only English tweets were considered based on the language identified by Twitter. Table I states the number of tweets collected for each event, and the number of distinct users who posted them.

Identifying communal tweets: Tweets posted during disaster events include both situational tweets (which contribute to situational awareness) and non-situational tweets. Since communal tweets are likely to be included among non-situational tweets, we first separated out non-situational tweets from situational ones, using the classifier proposed in our prior work [10]. Thereafter, we identified communal tweets from among the non-situational tweets, using the methodology proposed by Burnap *et al* [1], [8]. In brief, a classifier is developed based on the presence of specific hate terms in the content of the tweet, to decide whether a tweet contains abuse / hate towards some religious community.

Henceforth, we refer to the tweets which were categorized as communal by the classifier as *communal tweets*, and the users who posted these tweets as *communal users*. Table II gives some examples of communal tweets identified. We characterize the communal tweets and users in the next section.

IV. Characterizing communal tweets and users who post them

In this section, we analyze the communal tweets and the users who posted them. Specifically, we compare the set of communal tweets and communal users during a particular event with an equal number of randomly sampled non-communal tweets (as judged by the above classifier) and the users who posted them (referred to as *non-communal users*) during the same event.



(a) NEQuake (b) CShoot

Fig. 1: CDF of retweet-count of communal and non-communal tweets. Communal tweets are retweeted more.

A. Characterizing communal tweets

Which communities are targeted? Table II gives some examples of communal tweets and also identifies the communities targeted through those tweets. We find that the social / religious communities that are targeted, varies from one disaster to another. During man-made disasters, like terrorist attacks, the targeted community is most often the community to which the attackers are affiliated. However, some other communities are also targeted; for instance, during the Paris attack (PAttack), Christians were also targeted along with Islamic people. Surprisingly, even during natural disasters like NEQuake or KFlood, communal tweets get posted targeting some religious communities like Islamic people and Christian missionaries.

Popularity of communal tweets: We next investigate whether communal tweets become popular or receive large exposure among the population. For this, we check the retweet-count of a tweet, which is a standard metric for the popularity of a tweet and the exposure it receives.[1] Figure 1 shows the distribution of retweet-counts for communal and non-communal tweets, for the two events NEQuake and CShoot. Note that for this analysis, we only considered original tweets, i.e., tweets which are not retweets themselves. It is evident that, in general, *communal tweets are retweeted more than non-communal tweets.* A similar pattern was observed for the other events (omitted for brevity).

B. Characterization of communal users

We next analyze the users who post communal tweets during the disaster events. For this, we divide the users who post communal tweets into two categories – (i) **Originators:** users who originally post communal tweets, and (ii) **Propagators:** users who retweet the communal tweets posted by originators or some other propagators.

We next study the properties of originators and propagators separately.

Popularity of originators and propagators: We check whether communal tweets are posted by popular users or common masses. Across all the five different events, a similar

[1] We re-crawled all the tweets after several months from the date of the events, and hence collected the final retweet-count of the tweets.

TABLE II: Communities targeted during disaster events.

Event	Communities Targeted	Sample communal tweets
KFlood	Muslims	*Muslim* pigs wil never appreciate our Rescue forces of #India ! I ask them 2 selectively pick up ppl n let psychos drown #kashmirFloods
NEQuake	Christians	Look at the disgusting mentality of these *Christian* Missionaries. #NepalEarthquake [url]
	Muslims	#NepalEarthquake: No *Muslim* died Allah's Miracle!!!!!!! Lulzzzzzzzz [url] [url]
PAttack	Christians	If u think bringing any "persecuted Christians" into America from Syria and no terrorists will slip through, you're a f**ing idiot. #Paris
	Muslims	Apparently these awful Paris attacks were carried out by *Islamic* radicals. I'm as shocked as anyone who hasn't watched the news since 1985.

TABLE III: Sample users posted communal tweets.

Role played	Screen name	Follower count	Listed count
Originator	abhijitmajumder	69,532	750
	KiranKS	60,110	611
Propagator	SanghParivarOrg	133,922	464
	bjpsamvad	109,665	191

phenomenon is observed — communal tweets are posted and propagated by both common masses (25% having less than 100 followers) as well as by popular users (10% having more than 10,000 followers). Especially, some originators and propagators of communal tweets have several tens or hundreds of thousands of followers. Table III provides examples of some such popular communal originators and propagators.

Do originators also work as propagators? Next, we check whether originators of communal tweets during a disaster event also work as propagators during the same event. For this, we computed the *Szymkiewicz-Simpson similarity* score between the set of originators and the set of propagators during each event, and averaged the score obtained from the five different events. We find a low similarity score of 0.12. Thus, originators of communal tweets are mainly interested in posting their own opinion and thoughts, rather than retweeting contents posted by others.

Also, interestingly, the overlap between originators and propagators of *non-communal* tweets is two times higher than that for communal originators and propagators.

User overlap across different events: We also analyzed whether there exists a common set of users who originate / propagate communal tweets during multiple events. We found that across different events which occurred in the same geographical region (e.g., NEQuake, KFlood, GShoot, all of which occurred in the Indian subcontinent), there is a small set of common users who post communal tweets across all the events. For instance, originators like 'RamraoKP_', 'simbamara' and propagators like 'IndiaAnalyst', 'HinduRajyam' posted communal tweets during all these three events. However, in general, there is low overlap (about 8%) among the users who post communal tweets during different events.

Topical interests of communal users: We next attempt to identify the topical interests of communal users, over a set of broad topics: (i) Media & Journalism, (ii) Politics, (iii) Entertainment, (iv) Religion, (v) Sports, (vi) Writers/Authors and (vii) Business. Certain keywords which characterize these broad topics were collected through various online sources, such as *http://www.studentnewsdaily.com/media-vocabulary/* and *https://www.keywordspy.com/Category/Sports*.

TABLE IV: Distribution of topics of interest of common and popular originators of communal tweets.

User	Broad topic of interest						
	Media	Politics	Religion	Sports	Entertainment	Writing	Business
Popular users	47%	39%	5%	2%	4%	2%	1%
Common users	21%	29%	19%	14%	9%	6%	2%

For this analysis, we divide the users into two categories – (i) common users, having less than 5,000 followers, and (ii) popular users having $\geq$ 10,000 followers. For the common users, we use their Twitter account bio to identify their interests, by checking whether their bio contains the keywords corresponding to any of the broad topics stated above. For the popular users, along with checking their bio, we also used the methodology of our prior work [11] which identified the topical characteristics of millions of popular Twitter users, and then checked if the inferred topical characteristics match the keywords corresponding to any of the broad topics.

Table IV shows the distribution of topical interests of popular and common originators. A similar phenomena is also observed for propagators. For popular originators, most of them are related to media houses and politics. For the common users, a significant fraction of users are also interested in religion and sports, along with news media and politics.

Further, we checked the most frequent words which appear in the account bio and the tweets posted by the communal users and non-communal users. The account bios and tweets were pre-processed using standard techniques such as case-folding, removal of a common set of stopwords, and so on. Table V shows the top 5 words which appear in the account bio of the users for each category, and the tweets posted by these users. As expected, the communal users are mostly described by words related to religion and politics. On the other hand, the non-communal users are mostly described by words related to day-to-day conversation and positive sentimental words such as 'lover', 'life', 'fan' and so on.

C. Interactions among the users

We now investigate how the communal and non-communal users interact among themselves. In Twitter, the primary ways by which a user u can interact with another user v are (i) u can subscribe to the content posted by v by following v (ii) u can

TABLE V: Comparing the profile bio and tweets posted by users who posted communal tweets, and other users.

Most frequent words in bio	
communal	hindu, india, life, religion, endorsement
non-communal	indian, lover, fan, music, life
Most frequent words in tweets	
communal	modi, media, hindu, congress, muslim
non-communal	india, people, modi, bjp, govt

TABLE VI: Reciprocity and density of the mention and subscription networks among different groups of users.

Event	User group	Mention Network		Subscription Network	
		Reciprocity	Density	Reciprocity	Density
NEQuake	communal	4.53%	0.0032	24.94%	0.0095
	non communal	1.81%	0.0001	18.04%	0.0002
GShoot	communal	5.03%	0.0048	26.65%	0.0153
	non communal	1.45%	0.0002	18%	0.0018

@mention v in her tweet. We construct two types of interaction networks among the users, based upon these two modes of interaction. The first is a *friend network* where a directed link $u \rightarrow v$ indicates that user (node) u subscribes to the content posted by user v. The second is a *mention network* where the link $u \rightarrow v$ indicates that user u has @mentioned v.

To quantify the level of interaction among the users, we measure two structural properties of the subscription and mention networks – (i) *density*, which measures what fraction of all links that can be present in a network, are actually present, and (ii) *reciprocity*, which measures what fraction of the directed links are reciprocated, i.e., both the links $u \rightarrow v$ and $v \rightarrow u$ exist in the network. The importance of reciprocity is that if two users share a reciprocal link, then the two users are *mutual friends* with a higher probability (as compared to the chance of a fan subscribing to a celebrity, but the celebrity not reciprocating).

Table VI shows the reciprocity and density of the mention and subscription networks among different groups of users. We observed similar trend across all the disaster events; here we report the result for two disaster events — NEQuake and GShoot. We find that, for both the subscription network and the mention network, both the density and the reciprocity are significantly higher for the communal users compared to that for the non-communal users. These results indicate that a much larger fraction of the communal users are mutual friends, as compared to the non-communal users. Thus, the communal users largely interact among themselves, and form strongly-tied communities in the social network.

V. Concluding discussion

To our knowledge, this is the first attempt to characterize communal tweets posted during disaster events, and the users who post such tweets. We found that communal tweets are posted by many popular users who are mostly interested in politics and the media, and who form strongly tied communities in the Twitter social network. Also, communal tweets are

TABLE VII: Examples of anti-communal tweets posted during disasters.

Sad commentary of our times that people bring religion even into the devastating #NepalEarthquake
Won't Hindus remember Shiva or Hanuman in this crisis? So what's wrong if Christians remember Jesus? #NepalEarthquake
Saddened to know abt #GurdaspurAttack, pls avoid politics on terror. Act fast, whatever be religion of terrorists.

retweeted heavily, which makes it necessary to counter the potential adverse effects of such tweets.

We end by indicating a potential way of countering such communal tweets. We observe that, during a disaster, while many people post communal tweets, there are some users who post *anti-communal* content, asking people to stop spreading communal posts. Table VII shows some examples of anti-communal tweets posted during some of the disaster events considered in this work. However, we find that such tweets receive much lesser exposure (retweets) compared to communal tweets. Our future work would be directed towards identifying such anti-communal posts, and facilitating the popularity of such posts in order to counter the adverse effects of communal tweets.

Acknowledgement: This research was partially supported by a grant from the Information Technology Research Academy (ITRA), DeITY, Government of India (Ref. No.: ITRA/15 (58)/ Mobile/ DISARM/ 05) Additionally, K. Rudra was supported by a fellowship from Tata Consultancy Services.

References

[1] P. Burnap and M. L. Williams, "Cyber hate speech on twitter: An application of machine classification and statistical modeling for policy and decision making," *Policy & Internet*, vol. 7, pp. 223–242, 2015.
[2] I. Chaudhry, "# hashtagging hate: Using twitter to track racism online," *First Monday*, vol. 20, no. 2, 2015.
[3] L. Silva, M. Mondal, D. Correa, F. Benevenuto, and I. Weber, "Analyzing the targets of hate in online social media," in *Proc. AAAI ICWSM*, 2016.
[4] E. Greevy and A. F. Smeaton, "Classifying racist texts using a support vector machine," in *Proc. ACM SIGIR*, 2004.
[5] K. Dinakar, B. Jones, C. Havasi, H. Lieberman, and R. Picard, "Common sense reasoning for detection, prevention, and mitigation of cyberbullying," *ACM Transactions on Interactive Intelligent Systems (TiiS)*, vol. 2, no. 3, p. 18, 2012.
[6] Y. Chen, Y. Zhou, S. Zhu, and H. Xu, "Detecting offensive language in social media to protect adolescent online safety," in *Proc. PASSAT Workshop (with IEEE SocialCom)*, 2012.
[7] A. Mahmud, K. Z. Ahmed, and M. Khan, "Detecting flames and insults in text," 2008.
[8] 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.
[9] "REST API Resources, Twitter Developers," 2015, https://dev.twitter.com/docs/api.
[10] K. Rudra, S. Ghosh, N. Ganguly, P. Goyal, and S. Ghosh, "Extracting situational information from microblogs during disaster events: a classification-summarization approach," in *Proc. CIKM*, 2015.
[11] N. Sharma, S. Ghosh, F. Benevenuto, N. Ganguly, and K. Gummadi, "Inferring Who-is-Who in the Twitter Social Network," in *Proc. WOSN Workshop*, 2012.