Abhijnan Chakraborty *Max Planck Institute for Software Systems, Germany*
Niloy Ganguly *Indian Institute of Technology Kharagpur, India*

Editors: Roy Choudhury and Haitham Hassanieh

ONLINE SOCIAL NETWORKS TO FOSTER LONG-TERM WELFARE

The rise of Online Social Networks (OSNs) and their growing complexities have had dramatic effects in real life: due to their popularity and ubiquitous coverage, they are a major tool today to spread crucial news and vital information across remote areas of the world. Such increasing availability has unearthed a large volume of user-generated content, consisting of user opinion as well as factual content that often does not appear in mainstream media, but captures the persistent attention of a huge section of the population.

On the other hand, the flip side is that neither the reliability of the content nor the trustworthiness/expertise of their sources are guaranteed. Thus, it is hard to differentiate real news from fake news, or facts from opinions. Furthermore, the current easy accessibility of user-level information has left the online systems more vulnerable to privacy breach [1], and more susceptible to manipulation of the underlying social systems. For instance, today OSNs are particularly leveraged to form opinions about various events, where the discussions are often steered by a section of the users having certain agendas, instead of the spontaneous outbursts of unassuming users [2].

Our research in the last few years has been directed towards systematically tackling several of these real-life problems, which can be categorized into two broad directions: (a) proposing methods that promote proper usage of social networks and exploit available information for human good, and (b) developing techniques to detect and decelerate potential misuses. The next two sections elaborate these two directions.

PROMOTING PROPER USE OF ONLINE SOCIAL NETWORKS

In this section, we discuss two key works that contribute to utilizing social media for human good. In particular, the first problem enriches user experience in online media, thereby helping users to gain more knowledge, more information and vital news in a timely manner. The second problem, on the other hand, allows social media to be used as a powerful tool to complement disaster management efforts.

Search and recommendation systems over OSNs

As the amount of information generated in OSNs explodes, users need to rely on search and recommendation systems to find important events and breaking news stories. However, building such systems over OSNs is a challenging task, especially since there is no easy way to establish the credibility of the authors of different posts. In social media, messages can be posted by global news organizations and celebrities to locally popular community organizers or activists, and from expert computer scientists or astrophysicists to spammers faking identities of well-known users. Thus, to ensure high quality search and recommendation outputs (and filter out dubious posts), it is critical to accurately infer the topical interests of different users, and to identify the experts posting on specific topics.

The limited size of social media posts and the prevalence of informal (often code-mixed) languages make the problem even more challenging, since the traditional text analysis approaches (e.g., topic modeling tool LDA [3]) do not work well on OSNs. To combat such issues, we proposed a novel crowdsourcing-based methodology to infer topical expertise that relies on how other users describe a particular user [4]. Specifically in Twitter, users can organize the accounts they follow into different "Lists" and, when creating a list, they typically provide a list name and optionally a list description. For example, Figure 1 shows a list named "Politics" created by a user to put politicians and political journalists in it. Such names (and descriptions) act as crowd annotations describing the expertise of the list members. We gathered such annotations in large scale and combined them to infer the topical expertise of different users. We demonstrated that the proposed methodologies are far more accurate than content-based techniques, in inferring a wide range of topical interest of users, and identifying topical experts [5, 6]. The endeavor has also resulted in the development of several Web-based systems on the Twitter platform, e.g., topical search systems: "Whom To Follow" (https://twitter-app.mpi-sws.org/whom-to-follow) and "What Is Happening" (https://twitter-app.mpi-sws.org/what-is-happening), a system for inferring topical interest/expertise of users: "Who Is Who" (https://twitter-app.mpi-sws.org/who-is-who) and "Who Likes What" (https://twitter-app.mpi-sws.org/who-likes-what).

Efficient utilization of OSNs during disasters

During natural or man-made disasters, in addition to conversational posts, important real-time information is also posted on OSNs [7]. However, as shown in Figure 2, the valuable *situational information* [8] (which provides updates about current disaster situations) is often immersed among hundreds of thousands of tweets, mostly containing the sentiment and opinion of the masses. To effectively utilize OSNs during disaster events, it is necessary to (i) extract the situational information from the large amounts of sentiment and opinion, and (ii) summarize the situational information in real-time, to help in decision-making when time is critical. When analyzing thousands of tweets posted during disasters, we observed that *content words*, such as nouns, numerals, locations, and verbs, provide key information about a situation. Thus, we developed summarization framework attempts to include more situational information by maximizing the coverage of content words [9]. Moreover, certain numerical information, such as the number of casualties, vary rapidly with time. We also devised a scheme to identify the objects of disaster-specific verbs (e.g., "kill" or "injure") to continuously update important time-varying actionable items, such as the number of casualties [9].

Furthermore, large volumes of situational tweet streams are scattered across various humanitarian categories, such as "infrastructure damage" or "missing or found people," etc. Our interactions with disaster aid and relief organizations revealed that having a distinct humanitarian category-based summary is often preferred over overall high-level summaries. We observed that each of these humanitarian categories contain information about various small scale sub-events, such as "airport shut" or "building collapse," etc. that essentially is a noun-verb pair. We thus developed a noun and verb pair–based method to detect sub-events which are more explainable, compared to a random collection of words related to disaster events [10].

FIGURE 1. Twitter users can organize other users in "Lists" to easily read their posts. The names of these lists act as social annotations to describe the expertise of the list members.

PREVENTING MISUSE OF ONLINE SOCIAL NETWORKS

Compared to the traditional mass media, absence of any editorial control in social media has led to the emergence of orchestrated campaigns, often run by questionable organizations, authoritarian governments and terrorist groups, who use social networks to influence people's opinions. For example, anti-vaccine movement supporters spread misinformation through OSNs, filling parents with needless fears about immunizing their children, etc. Overall, OSNs being the primary (and often only) source of news for a large fraction of people has resulted in multiple consequences: proliferation of online (often social media only) news outlets without any gatekeeping policies, cut-throat competition for user attention, unchecked propagation of misinformation, extremely biased opinion, and so on. We have attempted to design methods to prevent such misuse of OSNs.

Mitigating spam propagation in OSNs

We observed that spammers gain importance inside OSNs with the help of unsuspecting users who are open to support spammers. We reported that a vast majority

ayyo! not again! :(Blasts in Hyderabad, 7 Killed: TV REPORTS
oh no !! unconfirmed reports that the incident in #newtown #ct may be a school shooting. police on the way
58 dead, over 58,000 trapped as rain batters Uttarakhand, UP.....may god save d rest....NO RAIN is a problem....RAIN is a bigger problem
@IvanCabreraTV: #Hagupit is forecast to be @ Super Typhoon strength as it nears Philippines. [url] Oh no! Not again!

FIGURE 2. Examples of tweets containing multiple fragments, some of which (in black) convey situational information, while other fragments (in blue) are conversational in nature.

FIGURE 3. A clickbait tweet posted by BuzzFeed to attract users to visit the corresponding webpage.

of spam supporters were benign users themselves, but willing to follow back almost everyone (to gain popularity in the network), thus, in a way, legalizing spammers [11]. To combat such link farming on Twitter, we proposed a ranking algorithm CollusionRank, which pushes a spammer down in potential recommendation lists, thereby reducing its visibility [11]. Such an algorithm is a modified version of PageRank; here, the importance of a node is computed based on its nature of followings, instead of its followers. The concept of link farming over OSNs and the algorithm CollusionRank has been applied widely in other scenarios as well.

Detecting and preventing the proliferation of clickbaits

Compared to the offline media, in which the readers' allegiance to a particular newspaper were almost static, online media offers the readers a gamut of options. Moreover, most of the online media sites do not have any subscription charges and their revenue comes mainly from advertisements on their web pages. Essentially, every media outlet has to compete with many such outlets for reader attention and make their money from the clicks made by the readers. Therefore, to attract the readers to visit the media site and click on an article, they employ various techniques, such as coming up with catchy headlines accompanying the article links, which lure the readers to click on the links.

Such headlines are known as clickbait: the headlines provide forward-referencing cues to generate enough curiosity so that readers feel compelled to click on the link to fill in the knowledge gap. Figure 3 shows a tweet posted by BuzzFeed to lure users to click on the corresponding URL. While these baits may trick them into clicking, in the long-run, clickbaits don't usually live up to the expectation of the readers, and leave them disappointed. We developed the first automated classifier to detect whether a headline is clickbait or not [12]. Then, we explored ways to block certain clickbaits from appearing in different social media sites. Multiple user surveys and a detailed investigation of clickbait consumption on Twitter suggested that clickbaits that different readers like to block vary greatly, driven by a particular reader's interests [13]. Hence, instead of a generalized solution, we developed personalized classifiers for individual readers, which predict whether a reader would like to block a particular clickbait given her earlier block and click history [12]. To curb the proliferation of clickbaits in reality, we, and subsequently other researchers, have built browser extensions [12, 14], which can warn the readers about the possibility of being baited by clickbait headlines, and offer an option to block certain types of clickbaits they would not like to see during future encounters. These extensions have been used by thousands of OSN users worldwide.

Utilizing users' truth perceptions to aid in fact-checking

OSNs have been severely criticized by policy makers and media watchdog groups for allowing fake news stories to spread unchecked on their platforms [15]. To overcome the limitations of automatic fake news detection techniques [16], OSNs today rely on (human) experts at fact-checking organizations, such as Snopes and PolitiFact to fact-check questionable stories. Since this approach does not scale, OSNs need to select a subset of the stories for fact-checking. To select such stories, OSNs are encouraging their users to report any news story that they perceive as fake. Stories reported as fake by a large number of users are then prioritized for fact checking. Thus, OSNs today are relying on their *users' perceptions of the truthfulness of news stories to select stories* to fact-check.

In our first attempt at understanding how users perceive truth in news stories, and how biases in their perceptions might affect detecting and labeling fake news stories [17], we designed a novel test for users to *rapidly* assess how truthful or untruthful the claims in a news story are, and collected such perceptions of hundreds of users. We found that (i) for many stories, the average truth perception of the users differs significantly from the actual truth of the story, i.e., wisdom of crowd is inaccurate, (ii) across different stories, we find evidence for both false positive perception bias (i.e., a gullible user perceiving the story to be more true than it is in reality) and false negative perception bias (i.e., a cynical user perceiving a story to be more false than it is in reality), and (iii) users' political ideologies influence their truth perceptions for controversial stories. We plan to take lessons from these observations and design efficient mechanisms to utilize user perceptions while prioritizing stories for fact-checking.

Ensuring fairness in OSN recommendations

Crowdsourced recommendations, called *trending topics*, are important tools in finding important events and breaking news stories. Typically hashtags and key-phrases are recommended as trending when their usage by the crowds suddenly jumps at a particular time. Once a topic is selected as trending, it gets prominently displayed on the social media homepages, thus reaching a large user population. By analyzing the promoters of trending topics on Twitter, we found that the majority of Twitter trends are promoted by crowds with demographics that differ significantly from Twitter's overall user population, and certain demographic groups (e.g., middle-aged black females) are severely under-represented in the process [18]. To make the demographic biases of Twitter trends transparent, we developed and deployed a service, "Who Makes Twitter Trends" (https://twitter-app.mpi-sws.org/who-makes-trends), where one can find the trending topics in the States and check their promoter (and adopter) demographics.

Apart from demographic bias, the crowd often consists of actors like bots, spammers, or people running orchestrated campaigns. As crowdsourced algorithms are driven by data generated by them, their outputs will reflect the biases in the composition of the crowds. We proposed to reimagine crowdsourced recommendations as the outcomes of a *multi-winner election* that is periodically repeated, and then the observed biases in recommendations can be attributed to the unfairness in the electoral system. Since (a) *a vast majority of voters are silent*, (b) some voters may vote multiple times, and (c) voters choose from several thousands of potential candidates (topics or hashtags), today's trending topic(s) election algorithms are vulnerable to electing fringe and extremist trends with as low as 0.001% of the electorate support. To *fairly aggregate the preferences of all users* while recommending in OSNs, we borrowed ideas from prior research on social choice theory, and identified a voting mechanism, Single Transferable Vote (STV), as featuring many of the fairness properties that we desire. We developed an innovative mechanism to attribute preferences of the silent majority, which also makes STV completely operational. The proposed approach provides maximum user satisfaction, reduces demographic bias and cuts down drastically on topics disliked by most but hyperactively promoted by a few users [19].

CONCLUSION

In this paper, we have briefly discussed some of our research works on online social networks. As their popularity continues to increase, the nature of the posted contents and the mediums themselves are undergoing massive changes. For instance, pictures and videos are becoming the primary content being shared, replacing the dominance of text messages. At the same time, end-to-end encrypted WhatsApp or other messaging platforms are replacing other more easily analyzable public platforms like Twitter. With this changed landscape, multiple researchers (including ourselves) are constantly engaged in attempting to make such networks truly useful regardless of the type of content or the underlying technology [20]. Interested readers are strongly encouraged to check the proceedings of conferences relevant to the broad area of social computing, such as ACM Conference on Computer-Supported Cooperative Work and Social Computing (CSCW), AAAI Conference on Web And Social Media (ICWSM), etc., to check the most recent advances in this field. ■

Abhijnan Chakraborty is a post-doctoral researcher at the Max Planck Institute for Software Systems, Germany. He obtained his PhD from the Indian Institute of Technology Kharagpur, India. His research interests span the area of social computing and fairness in algorithmic decision making.

Niloy Ganguly is a professor in the Department of Computer Science and Engineering at the Indian Institute of Technology Kharagpur, India. He obtained his Btech from IIT Kharagpur and his Phd from IIEST, Shibpur. His current research encompasses social computing, machine learning, and network science.

REFERENCES

[1] J. Isaak, and M. J. Hanna. (2018). User data privacy: Facebook, Cambridge Analytica, and privacy protection. *Computer*, 51(8).

[2] R. Recuero, and R. Araujo. (2012). On the rise of artificial trending topics in Twitter. *ACM Hypertext*, 305-06.

[3] D.M. Blei, A. Y. Ng, and M. I. Jordan. (2003). Latent dirichlet allocation. In *JMLR*.

[4] S. Ghosh, N. Sharma, F. Benevenuto, N. Ganguly, and K. Gummadi. Cognos: Crowdsourcing search for topic experts in microblogs. (2012). In *ACM SIGIR*, 2012.

[5] M. B. Zafar, P. Bhattacharya, N. Ganguly, K. P. Gummadi, and S. Ghosh. (2015). Sampling content from online social networks: Comparing random vs. expert sampling of the Twitter stream. In *ACM TWEB*, 9(3).

[6] M. B. Zafar, P. Bhattacharya, N. Ganguly, S. Ghosh, and K. P. Gummadi. (2016). On the wisdom of experts vs. crowds: Discovering trustworthy topical news in microblogs. In *ACM CSCW*.

[7] M. A. Cameron, R. Power, B. Robinson, and J. Yin. (2002). Emergency situation awareness from Twitter for crisis management. In *WWW*.

[8] S. Verma, S. Vieweg, W. J. Corvey, L. Palen, J. H. Martin, M. Palmer, A. Schram, and K. M. Anderson. (2011). Natural language processing to the rescue? Extracting 'situational awareness' tweets during mass emergency. *AAAI ICWSM*.

[9] K. Rudra, S. Ghosh, N. Ganguly, P. Goyal, and S. Ghosh. Extracting situational information from microblogs during disaster events: A classification-summarization approach. In *ACM CIKM*, 2015.

[10] K. Rudra, P. Goyal, N. Ganguly, P. Mitra, and M. Imran. Identifying sub-events and summarizing disaster-related information from microblogs. In *ACM SIGIR*, 2018.

[11] S. Ghosh, B. Viswanath, F. Kooti, N. K. Sharma, G. Korlam, F. Benevenuto, N. Ganguly, and K. P. Gummadi. (2012). Understanding and combating link farming in the Twitter social network. *WWW*.

[12] A. Chakraborty, B. Paranjape, S. Kakarla, and N. Ganguly. (2016) Stop clickbait: Detecting and preventing clickbaits in online news media. In *IEEE/ACM ASONAM*, 2016.

[13] A. Chakraborty, R. Sarkar, A. Mrigen, and N. Ganguly. (2018). Tabloids in the era of social media? Understanding the production and consumption of clickbaits in Twitter. *ACM CSCW*.

[14] M.M.U. Rony, N. Hassan, and M. Yousuf. (2018). BaitBuster: A clickbait Identification Framework. In *AAAI*.

[15] H. Allcott, and M. Gentzkow. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2).

[16] L. Graves. Understanding the Promise and Limits of Automated Fact-checking. Factsheet, 2018.

[17] M. Babaei, A. Chakraborty, J. Kulshrestha, E. M. Redmiles, M. Cha and K. P. Gummadi. (2019). Analyzing biases in perception of truth in news stories and their implications for fact checking. In *ACM FAT**.

[18] A. Chakraborty, J. Messias, F. Benevenuto, S. Ghosh, N. Ganguly, and K. P. Gummadi. (2017). Who makes trends? Understanding demographic biases in crowdsourced recommendations. *AAAI ICWSM*.

[19] A. Chakraborty, G. K. Patro, N. Ganguly, and K. P. Gummadi. (2019). Equality of voice: towards fair representation in crowdsourced top-k recommendations. *ACM FAT**.

[20] G. Resende, P. Melo, H. Sousa, J. Messias, M. Vasconcelos, J. Almeida and F. Benevenuto. (2019). (Mis)information dissemination in WhatsApp: Gathering, analyzing and countermeasures. *WWW*.