Measuring User Credibility in Social Media

Mohammad-Ali Abbasi and Huan Liu

Computer Science and Engineering, Arizona State University
Ali.abbasi@asu.edu,Huan.liu@asu.edu

Abstract. People increasingly use social media to get first-hand news and information. During disasters such as Hurricane Sandy and the tsunami in Japan people used social media to report injuries as well as send out their requests. During social movements such as Occupy Wall Street (OWS) and the Arab Spring, people extensively used social media to organize their events and spread the news. As more people rely on social media for political, social, and business events, it is more susceptible to become a place for evildoers to use it to spread misinformation and rumors. Therefore, users have the challenge to discern which piece of information is credible or not. They also need to find ways to assess the credibility of information. This problem becomes more important when the source of the information is not known to the consumer.
In this paper we propose a method to measure user credibility in social media. We study the situations in which we cannot assess the credibility of the content or the credibility of the user (source of the information) based on the user's profile. We propose the *CredRank* algorithm to measure user credibility in social media. The algorithm analyzes social media users' online behavior to measure their credibility.

Keywords: Information Credibility, Behavior Analysis, Misinformation

1 Introduction

Using social media, people easily can communicate and publish whatever they like. As a result, people are able to create huge amounts of data. For example, users on Twitter create 340 million tweets every day. Users on YouTube upload 72 hours of video every minute. In wordpress.com alone, bloggers submit 500,000 new posts and these posts receive more than 400,000 comments everyday[1].

People use social media either for communications or share newsworthy information. They use social media for almost every aspect of their lives. They use social media during disasters to report injuries, damage, or their needs. Examples of events in which social media was utilized include the recent tsunami in Japan, Hurricanes Irene, and Sandy, and the earthquake in Haiti [2]. In business and marketing, social media is also used for product and service review and recommendation. As a result, it is very common to read reviews and comments

[1] http://www.jeffbullas.com/2012/08/02/blogging-statistics-facts-and-figures-in-2012-infographic/

before purchasing a product or using a service. During the social events such as the *Arab Spring* and the *Occupy Wall Street (OWS)* movement, social media was effective to spread information about the movements[13]. In many cases, people report incidents and events almost instantly and their reports cover different aspects of the event (people act as social sensors). Social media provides first-hand data, but one pressing problem is to distinguish true information from misinformation and rumors. In many cases, social media data is user generated and can be biased, inaccurate, and subjective. Furthermore, some people use social media to spread rumor and misinformation[2] . Consequently, information in social media is not necessarily of equal value, and we need to assess the credibility of the data before using it for decision making.

Using credible information is a prerequisite for accurate analysis utilizing social media data. Non-credible data will lead to inaccurate analysis, decision making and predictions. Credibility is defined as "the quality of being trustworthy". In communication research, information credibility has three parts, message credibility, source credibility, and media credibility [14]. Comparing conventional media, assessing information credibility in social media is the more challenging problem. In the case of conventional media such as newspapers, the source and media are known; in addition the medium's owners take responsibility for the content. However in the case of social media, the source can be unknown thus no one takes responsibility about the content. In many cases a *username* is the only information we have about its source (e.g., an incomplete or even fake profile in Twitter or YouTube that publishes information about an incident).

Ranking social media users on their credibility is one approach to measure the credibility of the given piece of information. Twitter, for example, has a set of verified accounts. These accounts have a blue badge on their profiles. According to Twitter, "The verified badge helps users discover high-quality sources of information and trust that a legitimate source is authoring the account's tweets."[3] Neither Twitter nor other social media websites are able and want to verify all their users. In the best scenario, only a small portion of the users can be verified by the websites. Considering this and the fact that many users would prefer to remain unknown, it is expected that the majority of users in social media are unverified.

This anonymity is both an advantage and a disadvantage of social media. On one hand, people create content, and leave feedback or vote without being afraid of any negative side effects resulting from their activities. This is a great advantage especially for people in countries that lack the freedom of speech. On the other hand, people could also take advantage of openness and anonymity. Some would create many accounts in which to leave positive reviews in order to boost one product or negative reviews to downgrade another. During social and political movements (e.g., Arab Spring revolutions), one could observe many highly organized Twitter accounts that actively tweet against the revolution[1].

[2] http://personal.stevens.edu/ ysakamot/726/paper/Grant/RAPIDdescription.pdf
[3] http://www.twitter.com/verified

We would see the same misbehavior in social bookmarking systems in which highly coordinated accounts would try to change the voting results in a specific direction. Such behavior could significantly decrease the quality of social media content.

Contributions We propose to address user credibility to tackle the information credibility problem in social media. We propose the *CredRank algorithm*, which measures the credibility of social media users based on their online behavior.

Paper organization The rest of the paper is organized as follows. Section 2, *Literature Review*, reviews the related work on information credibility and credibility in social media. Section 3, *Problem Statement*, discusses the credibility problem in social media. Section 4, *A Proposed Solution*, introduces the CredRank algorithm as one solution to measure user credibility. Section 5, *Experiments*, shows the use of the proposed method on the U.S. Senate voting record data. Section 6, *Discussion*, summarizes findings and describes future work.

2 Literature Review

Assessing credibility is an important part of research on mass communication. Seminal work on credibility concentrated on source credibility as well as credibility attributed to different media channels [9]. In traditional media as well as social media, the credibility of the source has a great effect on the process of acquiring the content and changing audience attitudes and beliefs [5]. Studies confirm that people consider Internet information as credible as traditional media such as television, radio, and magazines but not as credible as newspapers [11].

Castillo et al. in [6] discussed the information credibility of news propagated through Twitter. They used users' profile information, network information, and users' behavior (tweets and retweets) to assess the credibility of tweets. Barbier and Liu in [4] proposed a method to find provenance paths leading to sources of the information to evaluate its credibility. Researches have used trust information to evaluate online content. Trust is widely exploited to help online users collect reliable information in applications such as high-quality reviews detection and product recommendations [10]. Guha et al. [8] studied the problem of propagating trust and distrust among Epinions' [4] users, who may assign positive (trust) and negative (distrust) ratings to one another. They used trust information to rank users and using that, rank the content generated by the users. Castilo et al. [6], used features from users' posting behavior (tweet and retweet), text, and the network (# of friends and # of followers) to distinguish credible from not credible tweets. Agichtein et al. [3], used community information to identify high quality content in question and answering portal (Yahoo! Answers [5]). They used features from *answers*, *questions*, *votes*, and *users' information and relationship* to build a model to measure quality of the content.

[4] http://www.epinions.com
[5] http://answers.yahoo.com

Popularity is the most accepted measure of assessing credibility of users and content in social media. Usually popularity and credibility are used interchangeably. For example many users would trust a Twitter user who has many followers. Similarly one might trust a piece of information (a video clip on YouTube) if many people had already watched it. Using popularity idea, there are some work that used link based information (e.g., PageRank and HITS) to rank the users and evaluate the content based on the source's rank. [12] used HITS to rank users and find experts and high quality answers in the question and answering communities. Using number of inlinks (# of friends on Facebook or # of followers on Twitter) is well-accepted feature to measure the importance (credibility or influence in different concepts) or users. Cha et al. [7] use three approaches (*indegree*, *retweet*, and *mention*) to measure users' importance in Twitter. The study shows that although *indegree* measures the popularity of a user, it does not necessarily reflect the importance (or influence in some domains) of the user.

3 Non-credible Information and the Need for Detection

Non-credible users are responsible for part of the non-credible information in social media. In this section, using real examples from the Arab Spring movements in social media sites, we show how users can generate and spread misinformation or prevent the spread of trustworthy information.

Twitter Usually for disasters or social events, the most useful and novel information is distributed by unknown or unpopular social media users. In this situation, the receiver of the information does not have adequate time and/or resources to assess source's credibility. Since they cannot assess the credibility of the information or sources, ordinarily the user relies on the popularity of the source or the content (# of followers or # of retweets). However ordinary Twitter users do not have many followers; therefore, their content would not get attention and would be lost among many other tweets. On the other hand some organized users take advantage of this situation. For example, during Arab Spring there were many coordinated users tweeting against the revolutionists. Many of them could be government-supported accounts. These users were very organized and most of them followed one another and retweeted one another's tweets. Therefore, their content could easily be noticed in Twitter.

YouTube In some cases, coordinated users targeted a video clip on YouTube to take it down. They frequently submitted false reports against the video and in many cases, these false reports led to the video removal from YouTube.

Voting Systems In voting systems (social bookmarking systems), users who have more supporters can get more votes and publicize their content more easily than others. It is common in this kind of system that many users create a clique, usually vote for one another's content, and against other users' contents. This enables them to publicize their content and deter other users who cannot collect enough votes for their content. These highly connected users also easily can degrade others' content by awarding them negative votes.

In all of these cases *coordinated users* can easily suppress independent users and prevent their content from spreading in social media. They also would be able to spread misinformation. These *coordinated users* have highly correlated behavior (e.g. their tweets are very similar or their votes have similar patterns). The next section shows our attempt on detecting these users by monitoring their online behavior. These are non-credible users who generate and spread misinformation or prevent other users from spreading their content. The main properties of non-credible users are:

- Creating a large number of accounts and using the accounts to spread the word.
- Voting, regardless of content, for other users in their group. As the votes go for the user, not the content, the number of votes coming from the group members does not represent the quality of the content[6].

4 Proposed Solution

Our solution gives each independent person an equal vote and a chance to publicize his/her content. We perform the following steps: (1) *detect and cluster* coordinated users (dependent users) together and (2) *weight* each cluster based on the size of the cluster. We design the *CredRank* algorithm to perform these two steps.

4.1 CredRank Algorithm

This algorithm finds users with similar behavior and clusters them. CredRank uses a hierarchical clustering method to cluster similar users into clusters. We measure similarity of behaviors to calculate the similarity between users.

$$Sim(u_i, u_j) = \frac{1}{t_n - t_0} \sum_{t=t_0}^{t_n} \sigma(B(u_i, t), B(u_j, t)) \tag{1}$$

where $B(u_i, t)$ is user u_i's behavior in timestamp t and $\sigma(B(u_i, t), B(u_j, t))$ is a function that measures the similarity of two users' behavior in the given timestamp t.

For each domain we use a specific function to measure similarity. For example, to measure Twitter users' similarity we calculate the similarity of their tweets. In social bookmarking systems, we measure the similarity of their votes. We can use various similarity measure approaches such as *edit-distance*, *tf-idf*, or *Jaccard's coefficient*. In our experiments we use Jaccard's coefficient to calculate behaviors' similarity.

$$\sigma(B_i, B_j) = \frac{|B_i \cap B_j|}{|B_i \cup B_j|} \tag{2}$$

[6] This problem also exists in political parties. In many cases, regardless of the topic, legislators vote for their party

Algorithm 1 CredRank Algorithm

1: Measure the pairwise similarity between users based on their behavior ($Sim(u_i, u_j)$).
2: Cluster users together if their similarity exceeds the threshold τ.
3: Assign $\omega_{C_i} = \frac{\sqrt{|C_i|}}{\sum_j \sqrt{|C_j|}}$ to each cluster, which is the cluster's weight. Each member in the cluster C_i, has a weight of $\frac{\sqrt{|C_i|}}{|C_i|}$ which is the credibility assigned to the member.

$\sigma(B_i, B_j) = 1$ shows that two users' behaviors are completely similar and $\sigma(B_i, B_j) = 0$ shows that their behaviors are different.

After calculating the similarity between users, we cluster users together if their similarity exceeds the threshold τ. The value of τ varies for different domains.

In the next step, using the following formula, we assign clusters' weights.

$$\omega_{C_i} = \frac{\sqrt{|C_i|}}{\sum_j \sqrt{|C_j|}} \tag{3}$$

where ω_{C_i} is the weight assigned to the cluster C_i with $|C_i|$ members. Each member in cluster C_i, has a weight of $\frac{\sqrt{|C_i|}}{|C_i|}$. This value show the amount of the credibility associates with the member.

5 Experiments

We use US Senate voting history data to show how the proposed algorithm helps us to detect coordinated collective behavior. In this case we consider the highly coordinated voting as non-credible behaviors. In this section we show that how we can detect these coordinated behavior. Then we use these coordinated behaviors to detect senators with similar voting history. Then we cluster senators with similar voting history in the same groups. Our analysis show that usually votes in each cluster are highly correlated.

Dataset We crawled United States Senate official websites [7] to collect senators' votes records. The website provides Senate "Roll Call Vote" results for the current and several prior Congresses. We crawled voting history from 1989 to 2012. For each issue, we collected each Senator's vote (yea or nay) and the vote result (rejected, agreed to, passed, and confirmed).

To analyze the correlation between votes, we use CredRank algorithm idea and calculate top eigenvalues as we report them in table 1.

[7] http://www.senate.gov

Table 1: Top eigenvalues of senators' voting history

Eig 1	Eig 2	Eig 3	Eig 4	Eig 5	Eig 6
70.3542	12.2152	1.6815	0.9427	0.7385	0.7013

By using k-means, with different values of k, in most cases Democratic and Republican senators cluster into different clusters. In almost all cases senators' votes depend on their party. In only a few cases the votes really represent senator's own opinion. Oftentimes votes of a few independent senators are highly influential on the result of votes and usually the result is highly dependent on their votes.

Referring to Table 1, the results show that Senators' votes are highly correlated. By using CredRank we can cluster senators into 6 clusters and rank them based on the number of senators in each group. If we pick one representative for each group, with the weight calculated by step 2 of the algorithm, we would be able to generate the same vote results as of votes from all of the senators.

Despite the real-world voting systems, we do not expect to observe coordinated voting behavior in social media. If two users in a rating system have very similar votes for many products, we consider this voting behavior as non-credible behavior. Therefore the users considered as non-credible users.

6 Discussion

In this paper, we propose a method to detect coordinated behavior in social media and assign a lower credibility weight to users who are involved in the coordinated behavior. In this process, we are able to prevent the spread of misinformation generated by these users, which is an attempt to increase the quality of information in social media. The proposed algorithm helps us to detect individuals who use many social media accounts and do so in a way to diffuse their content. The CredRank algorithm can be used in many cases such as: preventing the distribution of rumors, averting coordinated activities, and thwarting fake product reviews. In the future work, we will improve the algorithm to solve two drawbacks we mention them next. Using the method might prevent true diffusion of information. In addition, calculating similarity among all users' behaviors in real time might be computationally expensive.

In order to achieve better results, we must consider all three parts involved in information credibility, including message credibility, source credibility, and media credibility. Focusing on source credibility and considering more features of sources, such as network and profile, to assess user credibility is an extension to this work. By considering behavior, network, and profile we expect to construct a reliable model to assess source credibility in social media.

Acknowledgments

This research is sponsored, in part, by Office of Naval Research (Grant number: N000141110527).

References

1. M. Abbasi, S. Chai, H. Liu, and K. Sagoo. Real-world behavior analysis through a social media lens. *Social Computing, Behavioral-Cultural Modeling and Prediction*, pages 18–26, 2012.
2. M. Abbasi, S. Kumar, J. Filho, and H. Liu. Lessons learned in using social media for disaster relief-asu crisis response game. *Social Computing, Behavioral-Cultural Modeling and Prediction*, pages 282–289, 2012.
3. E. Agichtein, C. Castillo, D. Donato, A. Gionis, and G. Mishne. Finding high-quality content in social media. In *Proceedings of the international conference on Web search and web data mining*, pages 183–194. ACM, 2008.
4. G. Barbier and H. Liu. Information provenance in social media. *Social Computing, Behavioral-Cultural Modeling and Prediction*, pages 276–283, 2011.
5. J. Burgoon and J. Hale. The fundamental topoi of relational communication. *Communication Monographs*, 51(3):193–214, 1984.
6. C. Castillo, M. Mendoza, and B. Poblete. Information credibility on twitter. In *Proceedings of the 20th international conference on World wide web*, pages 675–684. ACM, 2011.
7. M. Cha, H. Haddadi, F. Benevenuto, and K. Gummadi. Measuring user influence in twitter: The million follower fallacy. In *4th International AAAI Conference on Weblogs and Social Media (ICWSM)*, 2010.
8. R. Guha, R. Kumar, P. Raghavan, and A. Tomkins. Propagation of trust and distrust. In *Proceedings of the 13th international conference on World Wide Web*, pages 403–412. ACM, 2004.
9. C. Hovland and W. Weiss. The influence of source credibility on communication effectiveness. *Public opinion quarterly*, 15(4):635–650, 1951.
10. M. Jamali and M. Ester. Trustwalker: a random walk model for combining trust-based and item-based recommendation. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 397–406. ACM, 2009.
11. T. Johnson and B. Kaye. Cruising is believing?: Comparing internet and traditional sources on media credibility measures. *Journalism & Mass Communication Quarterly*, 75(2):325–340, 1998.
12. P. Jurczyk and E. Agichtein. Discovering authorities in question answer communities by using link analysis. In *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, pages 919–922. ACM, 2007.
13. H. Kwak, C. Lee, H. Park, and S. Moon. What is twitter, a social network or a news media? In *Proceedings of the 19th international conference on World wide web*, pages 591–600. ACM, 2010.
14. M. Metzger, A. Flanagin, K. Eyal, D. Lemus, and R. McCann. Credibility for the 21st century: Integrating perspectives on source, message, and media credibility in the contemporary media environment. *Communication yearbook*, 27:293–336, 2003.