A Fuzzy Model for Combating Misinformation in Social Network Twitter

Dubravka Gavric and Armen Bagdasaryan
Department of Mathematics, College of Engineering and Technology, American University of the Middle East, Kuwait

E-mail: dubravka.gavric@aum.edu.kw; armen.bagdasaryan@aum.edu.kw

Abstract. Constant and rapid increase of social media users implies the increase of spread of misinformation in social media networks. One of the leading networks Twitter is becoming significant source of information and news for online users. This research proposes a new approach to the problem of combating misinformation in social media network Twitter. The approach is based on the SI (Epidemic) Fuzzy Model to combat the spread of misinformation on Twitter. The mathematical model is given by a system of differential equations including fuzzy parameters and factors that describe various characteristics of misinformation spread in complex network. We also reflect on current challenges in combating misinformation in social media with the goal to stimulate future research in that domain by pointing out important factors that need to be taken into account while developing a model for combating misinformation.

1. Introduction
Twitter is a social network that allows users to share text messages. These messages are short and they are called tweets. Since 2006, millions of people around the world have been using it regularly. However, there is little accountability and source validation while different information is being progressively shared. Twitter and all other social media networks have no perfect tools to stop fake news. As a result there is uncontrolled flow of misinformation and unknown size (percentage) of misinformed users. This phenomenon is not possible to describe mathematically precise but it is possible to approximate certain data in order to have better information about misinformation spread and its final stage.

Twitter is broadcast medium where people exchange opinions, news, events that are happening around the world. It can be represented as graph in which the direction of links determines the flow of information. Users with many followers (large in-degrees) can effectively spread information to big number of other users (large number of nodes). The same is true when it comes to the spread of misinformation. Misinformation which spreads fast usually has short-term effect and long term consequences [1].

Misleading news are globally dangerous as it can be used to instigate hatred, conflicts, racism, theological intolerance, crime and general confusion.

In order to have better quality information environment on Twitter social network there is a need to develop a good strategy to either discredit the sources of misinformation or to stop it (decrease it) while it is already sent.

1 To whom any correspondence should be addressed.
This paper is organized in the following way. After this introductory section, the second section presents important facts about Twitter social network. Third section introduces challenges in combating misinformation on Twitter. Fourth section proposes fuzzy model for combating misinformation in social network Twitter. Finally, fifth section gives conclusions and stimulates further research on this subject.

2. Important facts about Twitter
Twitter has about 330 million monthly active users. It is social networking and micro-blogging service which enables users to read and post messages limited to 280 characters. Users can also share photos and short videos. The most influential users have over 100 million followers [2].

Twitter users are not like users in typical social networks. Twitter striking popularity has attracted many high-profile users such as celebrities, politicians, media representatives and other influential. Therefore the spread of misinformation which comes through influential users is much faster than the one which comes from ordinary users.

Twitter user can follow other user and the user being followed does not have to follow back. User which follows receives all messages (tweets) from the one he follows. Retweet is the message which can be further shared once received.

Twitter is powerful medium which can collect data from the network precise, quickly and with small cost. At Xerox Palo Alto Research Centre, its research manager Ed H Chi stated ‘Twitter is kind of this perfect laboratory for understanding how information spreads’ [3].

A quantitative analysis of tweets during Ebola crisis in 2014 reviles that misinformation on Twitter spreads just like true information. The SEIZ compartmental model is applied to information propagation on Twitter [4].

3. Challenges in combating misinformation on Twitter
Many researches are focused on finding information source which can be equally used for the misinformation source. In order to know misinformation diffusion one of the most important things is to discover its source. It is not always an easy task. It could be one source or many sources and based on that distribution changes.

Important observations are the state of the nodes and the timestamps at which nodes adopted the information. Tracing the source helps locate misinformation creator [5].

Another important point is to find the location of the beginning node(s). It has been shown that the influence of a node in certain situations is more dependent on its location than the number of connections it has [6]. This paper is about identifying influential people in a social network and uses SIS (susceptible-infected-susceptible), SIR (susceptible-infected-recovered) models. Immunization is limited to the node that is inoculated by external means and that aligns with the case where once a node is inoculated, it can inoculate more users (by virally spreading misinformation). Inoculation has a direct effect only on inoculated node, meaning that misinformation does not propagate [7].

The early detection of misinformation is important for its blockage. This was motivation for the research of stochastic TCMD problem which aim was to catch all possible cascades of misinformation within the given time span [8].

Some misinformation is highly interesting while other information might not be interesting to users and therefore its spreading will be less. There is a research in the field of psychology which is focused on sentimental analysis of tweets. The research [9] tracked national mood on Twitter, and found that people in America are much happier on Sunday morning than Thursday evening. It means that some information are differently received (with bigger impact degree) on Sunday morning than on Thursday evening. Information similar to this helps describing misinformation content impact on other users. It is important to take into account when trying to combat misinformation.

For example, The Web Ecology Project [10] concluded that content of the tweet is the best shared with news media users.

Another research [11] revealed that users who limit their tweet to a single topic showed the largest increase in their influence scores and that the content value of the tweet has impact on retweets. They observed 3 types of influence in Twitter social network:
- Indegree influence (the number of followers of a user)
Retweet influence (the number of retweets)
Mention influence (the number of mentions containing one's name)

Methodology for comparing user influence was by using statistics Spearman's rank correlation coefficient to measure the strength of the association between two rank sets (two different influence measures) in a dataset of N users. Measures like the number of tweets and out degree (the number of people a user follows) were not useful because they identified robots and spams as most influential. They found out that celebrities have more mention influence while news media have better retweet influence. Conclusion is that influence is not gained spontaneously but through constant efforts and personal involvement of user.

Influential Twitter users ranking and characteristics can be found in [12, 13].

To make a relevant modelling of influence in social media it is important to capture temporal dynamics (new influential users appearing over time and one users which disappear over time). This data are not precise and therefore the fuzzy modelling approach is very convenient [14, 15].

The novel algorithm based on fuzzy set theory is proposed for finding similar nodes in large directed graphs with millions of nodes and billions of connections. This algorithm is practically verified on Twitter social network case study [16]. Algorithm is provided with 163 representative influential Twitter user in the category of science and a total of 72 new users were discovered with the similarity measure above 0.5. Influential users are expected to propagate misinformation quickly and widely but one has to keep in mind that it is not always the case. There were examples when user activity was more important than its influence. In those occasions ordinary users shared misinformation faster than influential users. More information about Tweeter users and their behavior in different situations is available in [17].

Twitter has already experienced many spreads of misinformation on its network and there is a record of user behaviour towards certain type of misinformation. From those experiences can be given predictions for the future user behaviour in similar situations.

The research show that the number of misinformation is significantly bigger than fact checking on Twitter [18].

Another study on Twitter obtained true and false new stories from 2016-2017. It has been found out that misinformation is shared more than true information, especially the ones which contain false political information. The effectiveness of those misinformation holding tweets was larger than false news in the domain of science, finances etc. Also it is reported that misinformation spreads more than true information because humans, not robots, are likely to spread it. More precisely, according to the biggest study data into fake news on Twitter, the truth takes six times longer to be seen by 1500 people than misinformation [19].

The final stage of combating misinformation is to develop a model which can be verified. Most of the models is difficult or impossible to put in algorithm. Some researchers use the example of epidemic modelling in order to follow news and rumours on Twitter. In order to check the credibility of Twitter the Tweet Cred system is developed. It uses crowdworkers to annotate data and train machine learning algorithms that can learn from users after assessing the credibility of tweets [20]. Also there are tools to detect and follow misinformation on Twitter called Twitter trials [21]. Twitter trials do not monitor the stream automatically to detect misinformation but requires the user to input a rumour which desires to investigate. Therefore Floaxy platform is developed to automatically detect and analyse online misinformation.

Current limitations of the systems for combating misinformation are: providing alerts without rational explanation of their decision, treating users as passive consumers rather than active detectors of misinformation and the possibility that all automatic systems can be possibly fooled such that misinformation goes undetected.

4. Fuzzy model for combating misinformation on Twitter

There are huge similarities between the spread of misinformation in social networks and spread of epidemic diseases. Therefore the model that we propose is based on the ordinary differential equations which have applications in epidemiology. SI model origins from SIR model which was developed in 1927 by Kermack and McKendrik [22]. It is an epidemiological model that computes the number of individuals infected with a contagious disease in a closed population over time. The
SI Model is the simplest epidemic classical model to describe the directly transmitted diseases. The rates of birth and mortality, immunity and disease fatality rates are not included in the model. The SI model represents interaction between $S$ (susceptible) and $I$ (infected) individuals. Based on SI model, we propose a model to help tracking and combating misinformation on Twitter as follows.

The model is represented by the diagram:

$$S \xrightarrow{\beta} I$$

The differential equations are given by:

$$\frac{dI}{dt} = \beta SI$$

$$\frac{dS}{dt} = -\beta SI$$

where $S + I = 1$ represents all Twitter users.

- $S$ stands for Twitter users which did not receive misinformation.
- $I$ stands for Twitter users which received misinformation (The assumption is that they will share the tweet.).

The $I$ has impact on $S$. (The member of $S$ might become the member of $I$ as a consequence of shared misinformation)

The main characteristic of spread of any online information is that it spreads exponentially and quickly such that big number of users is affected in short period of time. Therefore the number of Twitter users which received misinformation at any instant time $t$ is given by:

$$I = \frac{I_0 e^{\beta t}}{S_0 + I_0 e^{\beta t}}$$

Fuzzy number $\beta$ is the sharing coefficient from $S$ to $I$.

$$\beta(v) = \begin{cases} 
0 & \text{if } v < v_{\min} \\
\frac{v - v_{\min}}{v_M - v_{\min}} & \text{if } v_{\min} \leq v \leq v_M \\
1 & \text{if } v_M < v_{\max} \end{cases}$$

The sharing coefficient $\beta = \beta(v)$ measures the chance of sharing tweet between $S$ and $I$ with an assigned influence power number $v$. Influence power is a number of different user’s characteristics which cause other users to believe to his shared tweet. So $\beta$ is a membership function of a fuzzy number. When the influence power number of the misinformation is very low the possibility of successful share of misinformation is insignificant. In that cases the misinformation will be combated. The information about user and its influence power one can obtain from Twitter REST API which provides programmatic access to read and write Twitter data [23]. The REST API identifies Twitter applications and users using OAuth; responses are available in JSON. It is possible to obtain profiles of Twitter users (profile contains basic information, as well as the number of followers and number of following), followers of certain user, and users that certain user is following. The base URI for all Twitter REST API calls is https://api.twitter.com/1.1.

- $v_{\min}$ is a minimum influence power needed for partially successful sharing of misinformation (means the other user will not completely believe to misinformation and will not share it to all).
• $v_M$ is the influence power needed for completely successful misinformation (means that the misinformation will be retweeted 100%).

• The maximum influence power number for every misinformation is limited by $v_{max}$. The main goal is to low $I$.

Here is the example how the influence power can be calculated. Suppose that the misinformation is about vaccination. An expert assigns the values for particular characteristics of influential user. The importance value can be calculated with some of the measures which are briefly classified in [13]. For this analysis the importance value numbers have been chosen arbitrarily and the chosen values do not reduce the generality of the conducted analysis.

Table 1. Each user characteristic is assigned the importance value.

| User’s characteristics          | Importance value |
|---------------------------------|------------------|
| Doctor                         | 2.2              |
| Many followers (ex. more than 1000) | 3.4           |
| Follows many (ex. more than 1000) | 1.1          |
| Female                         | 0.2              |
| Male                           | 0.3              |

Then the sum of all assigned values will be the influence power number $v$.

Among the group $I$ it is possible to select the most influential users and for this purpose we suggest the following.

We assume that every user has his influence power number $v$.

Table 2. Each user is assigned the influence power number.

| User | Influence power ($v$) |
|------|-----------------------|
| A    | 7.2                   |
| B    | 8.4                   |
| C    | 6.1                   |

After that we apply preference function $P: I \times I \to [0,1]$ where

• $P(A, B) = 0$ No preference to A (it does not mean that we prefer B)

• $P(A, B) \approx 0$ Slight preference to A

• $P(A, B) \approx 1$ Strong preference to A

• $P(A, B) = 1$ Complete preference to A

Let $d$ be the difference between influence powers of randomly selected two users from the set $I$. We will define a preference function $P(d)$ using $q$ and $p$ parameters where $q$ is the parameter of indifference and $p$ is the parameter of preference. We get $q$ by determining the difference between two numbers $v$ which is not important to us. For example, we can say that there is no significant difference between two users which have difference between their $v$ numbers less than 1. To get $p$, we can say that we absolutely prefer the first user in the pair if the difference between their $v$ values is more than 2.

Here is the preference function for the given example:

$$P(d) = \begin{cases} 
0 & \text{if } d \leq 1 \\
\frac{d-1}{2-1} & \text{if } 1 < d \leq 2 \\
1 & \text{if } d > 2 
\end{cases}$$

and in the general case
\[ P(\textbf{d}) = \begin{cases} 
0 & \text{if } d \leq q \\
\frac{d - q}{p - q} & \text{if } q < d \leq p \\
1 & \text{if } d > p
\end{cases} \]

It can be represented graphically as follows:

![Graphical representation of the preference function](image)

**Figure 1.** The preference function

After applying the preference function for all randomly selected pairs of users we get the following.

| Pairs | \(d\) | \(p(d)\) |
|-------|-------|--------|
| (A, B) | 7.2 – 8.4 = -1.2 ≤ 1 | 0 |
| (A, C) | 7.2 – 6.1 = 1.1 | 0.1 |
| (B, A) | 8.4 – 7.2 = 1.2 | 0.2 |
| (B, C) | 8.4 – 6.1 = 2.3 > 2 | 1 |
| (C, A) | 6.1 – 7.2 = -1.1 ≤ 1 | 0 |
| (C, B) | 6.1 – 8.4 = -2.3 ≤ 1 | 0 |

To find the most influential user we have to consider positive, negative and final preference.

| | Average 1 | Average 1– Average 2 |
|---|---|---|
| A | 0.05 | 0.05 – 0.2 = -0.15 |
| B | 0.6 | 0.6 – 0 = 0.6 |
| C | 0 | 0 – 0.55 = -0.55 |

Average 1 represents the positive preference, the Average 2 represents the negative preference and their difference gives us the final preference. In our case, the most influential Twitter user between users A, B and C is user B. It has the largest value as a result of the difference between the two averages.

### 5. Conclusion

In this paper we presented, to the best of our knowledge, the existing studies related to Twitter as well as the limitations of proposed methods for combating misinformation flow. We have described
challenges which are important to consider before developing a model to combat misinformation on Twitter. Finally we have offered a new fuzzy model which can be enriched in future researches.

6. References

[1] Pastor-Satorras R and Vespignani A 2001 Epidemic spreading in scale-free networks *Phys.Rev.Lett* 86 3200-03

[2] Twitter: Number of active users 2010-2019 *Statista* [Online] Available: https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users [Accessed: July 24, 2019]

[3] Savage N 2011 Twitter as Medium and Message *Communications Of The ACM* 54 18-20 Available: Business Source Complete, Ipswich, MA [Accessed: July 24, 2019]

[4] Jin F, Wang W and Ramakrishnan N 2014 Misinformation Propagation in the Age of Twitter *Computer* (00189162) [Online] 47 90 Available: Complementary Index Ipswich MA [Accessed: July 24, 2019]

[5] Zhu K and Ying L 2016 Information source detection in networks: Possibility and Impossibility results *The 35th Annual IEEE INFOCOM* 2016 pp 1-9

[6] Kitsak M, Gallos L K, Halvin S, Lijeros F, Muchnik L, Stanley H E and Makse H A 2010 Identifying influential spreaders in complex networks *Nature Physics* 6 888-893

[7] Meier D, Pignolet Y A, Schmid S, and Wattenhofer R 2014 On the windfall and price of friendship: Inoculation strategies on social networks *Computer Networks* 62 221-236

[8] Zhang H, Kuhnle A, Zhang H and Thai M T 2016 Detecting misinformation in online social networks before it is too late *Proc. Int. Conf. on Advances in Social Networks Analysis and Mining* (Davis) IEEE/ACM 541-548

[9] Fusaroli R, Perlman M, Mislove A, Paxton A, Matlock T and Dale R 2015 Timescales of Massive Human Entrainment *Plos ONE* [Online] 4 Available: Academic OneFile Ipswich MA [Accessed: July 24, 2019]

[10] Leavitt A, Burchard E, Fisher D and Gilbert S 2009 The influentials: New approaches for analyzing influence on Twitter *Web Ecology Project* 4 1-18

[11] Cha M, Haddadi H, Benevenuto F and Gummadi P K 2010 Measuring User Influence in Twitter: The Million Follower Fallacy *ICWSM* 10 10-17.

[12] Kanda Runapongsa S and Wit K 2014 Twitter influential users ranking using Twitter user characteristics *KKU Engineering Journal* 41 547-554

[13] Riqueime F and Gonzalez-Cantergiani P 2016 Measuring user influence on Twitter: A survey *Information Processing and Management* 5 949-975

[14] Sunitha M S and Mathew S 2013 Fuzzy graph theory: A survey *Ann. Pure Appl. Math.* 4 92-110

[15] Vascak J and Madarasz L 2010 Adaptation of fuzzy cognitive maps – a comparison study *Acta Polytechnica Hungarica* 7 109-122

[16] Jocic M, Pap E, Szakal A, Obradovic D and Konjevic Z 2017 Managing big data using fuzzy sets by directed graph node similarity *Acta Polytechnica Hungarica* 14 183-200

[17] Barabasi A L 2002 Linked (Cambridge MA: Persues Publishing)

[18] Shao C, Ciampaglia G L, Flammini A and Micerzer F 2016 Hoaxy: A platform for tracking online misinformation *Proc. of the 25th Int. Conf. Companion on World Wide Web* (Montreal) pp 745-750

[19] Vosoughi S, Roy D and Aral S 2018 The spread of true and false news online *Science* 359 1146-51

[20] Jin F, Dougherty E, Saraf P, Caao Y and Ramakrishnan N 2013 Epidemiological modeling of news and rumors on Twitter *Proc. of the 7th Workshop on Social Network Mining and Analysis* (Chicago) ACM 8

[21] Metaxas P T, Finn S and Mustafaraj E 2015 Using twittertrials.com to investigate rumor propagation *Proc.of the 18th ACM Conf. Companion on Computer Supported Cooperative Work & Social Computing* ACM pp 69-72
[22] Kermack W O and McKendrick A G 1927 A contribution to the mathematical theory of epidemics *Proc. Royal Society London* (A) 115 pp 700–721

[23] REST APIs *Twitter Developers* [Online] Available: https://dev.twitter.com/rest/public
[Accessed: July 24, 2019]