

# Online Social Network Analysis

## Marialaura Previti Second Year Ph.D. course activities

Tutor: Prof. Ing. Vincenza Carchiolo

Ph. D. Course in Systems, Energy, Computer and Telecommunications Engineering - XXXII Cycle

### ISSUE

Misinformation spreading over online social networks is becoming more and more critical due to the huge amount of information sources whose reliability is hard to establish; moreover, several human psychology factors as echo chambers and biased searches, plus the intensive use of bot, makes the scenario difficult to cope with. In this work, I investigate on the diffusion of true, false and mixed news through the Twitter network using a free large dataset of fact-checked rumor cascades.

My research goals are to assess how news spread depending on their veracity and to provide an analytic formulation of spreading process.

### DATASET

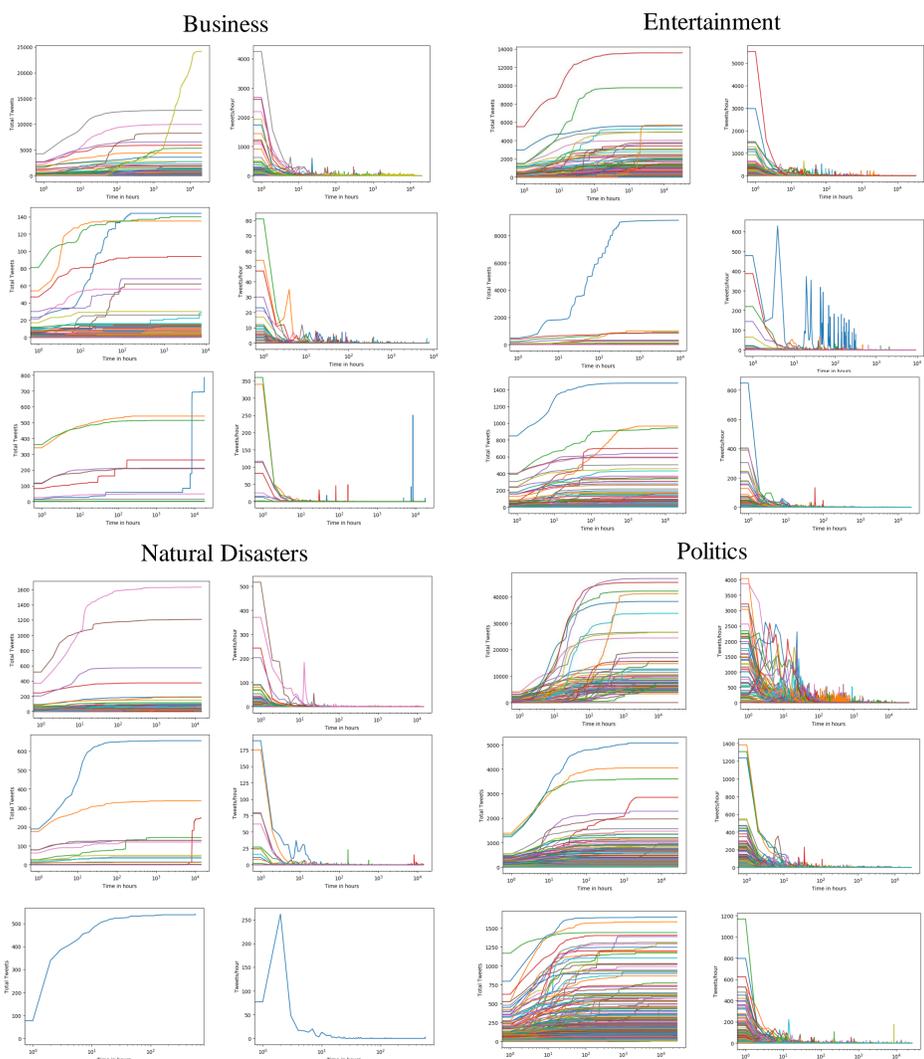
This work is based on part of the dataset used in [1]. The authors investigated the diffusion of true, false and mixed (i.e., partially true, partially false) news through the Twitter network using a dataset of fact-checked rumor cascades. Their data included **126,000 complete rumor cascades spread by 3 million people more than 4.5 million times.**

To assess news veracity or falsity, authors accessed to the full Twitter historical archives to collect all English language tweets that contained at least a link to one of six independent rumor debunking websites (snopes.com, politifact.com, factcheck.org, truthorfiction.com, hoax-slayer.com, and urbanlegends.about.com) from **September 2006 to December 2016**, then they reduced the number of tweets by taking only the ones that were replies to other tweets. For each reply tweet, they extracted the original tweet that they were replying to and then extracted all the retweets of the original tweet. To reconstruct the retweet graph, they used the **time-inferred diffusion method**, that compares the temporal order of the tweets with the follower graph in order to determine the correct path of each tweet over the network.

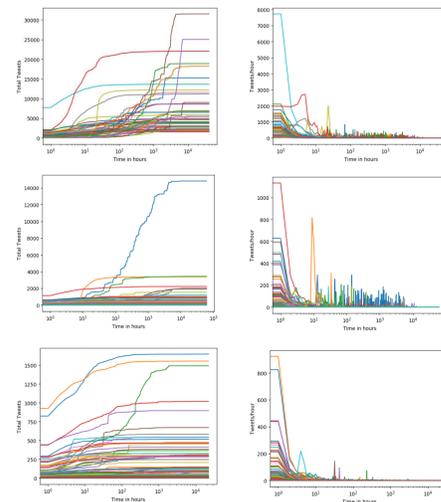
The fact-checking organizations behind the abovementioned websites, in addition to checking the news, tag them with a topic, so it is possible to divide them by subject areas. These tags have been reported for each tweet, so, thanks to a prefiltering activity, I divided the huge initial dataset into 7 smaller sized datasets each one representing a topic, I removed tweets without retweets and I rebuilt tweets cascades.

Topic	Number of tweets	Politics	False	Mixed	True
Politics	2094775	Tweets without retweets	10070	5323	5916
Urban Legends	801755	Total cascades	17076	1652	3441
Science and Technology	563291	<b>Urban legends</b>	<b>False</b>	<b>Mixed</b>	<b>True</b>
Entertainment	206926	Tweets without retweets	13177	3750	9390
Terrorism and war	205588	Total cascades	3281	899	1870
Business	151714	<b>Business</b>	<b>False</b>	<b>Mixed</b>	<b>True</b>
Natural Disasters	10687	Tweets without retweets	8167	2748	6
		Total cascades	2919	188	13
		<b>Science And Technology</b>	<b>False</b>	<b>Mixed</b>	<b>True</b>
		Tweets without retweets	9877	3107	47
		Total cascades	2056	529	126
		<b>Entertainment</b>	<b>False</b>	<b>Mixed</b>	<b>True</b>
		Tweets without retweets	4272	1	11
		Total cascades	1774	17	100
		<b>Terrorism And War</b>	<b>False</b>	<b>Mixed</b>	<b>True</b>
		Tweets without retweets	3768	276	2805
		Total cascades	4286	781	520
		<b>Business</b>	<b>False</b>	<b>Mixed</b>	<b>True</b>
		Tweets without retweets	8167	2748	6
		Total cascades	2919	188	13
		<b>Natural Disasters</b>	<b>False</b>	<b>Mixed</b>	<b>True</b>
		Tweets without retweets	1071	1	1
		Total cascades	247	15	1

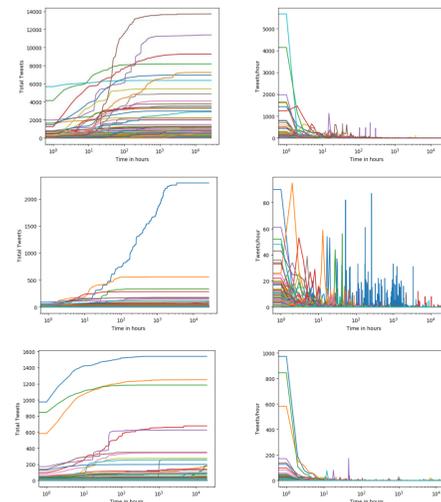
### Cascades plots: Total Tweets and Tweets/Hours (False-Mixed-True)



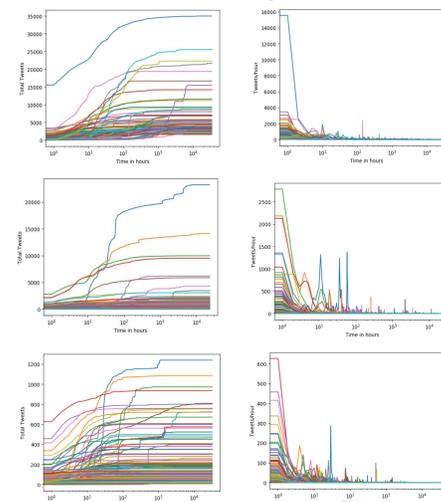
### Science and Technology



### Terrorism and War



### Urban Legends



### Retweet dynamic model

Observing the number of total tweets per cascade, I see that, despite the number of retweets is distinctly different (the total retweet curves of the false news have a higher slope and reach values much higher than other two graphs), all the curves are monotonic increasing with some steps in correspondence with the peaks present in the tweet per hour graphs. So it is possible to identify a parameterized function that approximates the value of each curve. To get a more regular trend, I calculated the average mean curve of the total tweets by taking the number of tweets per hour of each cascade and calculating the average point by point.

In order to model the evolution of retweets over time, I generated a model starting from the observation of our data subset characteristics. In the following, I indicate with  $R(t)$  the number of retweets at instant  $t$ . The number of retweets in a generic instant  $t + \Delta t$  will be given by the sum between the ones at instant  $t$  and the ones that will be posted in the following instants with a certain retweet rate  $\lambda$ , but not exceeding the number of total retweets of the curve  $N$ , hence:

$$R(t + \Delta t) = R(t) + \lambda[N - R(t)]\Delta t$$

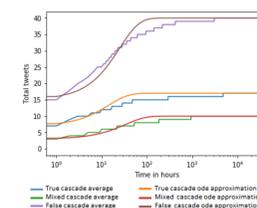
In the limit of a small  $dt$ , I have the following first-order ordinary differential equation:

$$R'(t) + \lambda R(t) = \lambda N$$

whose solution is:

$$R(t) = [R(t_0) - N]e^{-\lambda t} + N$$

where  $R(t_0)$  indicates the number of tweets at the initial instant.



By plotting this solution I obtained a set of curves with different slopes based on the  $\lambda$  value chosen. To find curves that best fit the original curves, I calculated the mean square errors (mse) between the curves with different  $\lambda$  values and the real data obtained from the dataset and returns the one with the lowest mse.

The curves that best fit the real values curves are:

- for true retweet cascades  $\lambda = 0,02$  with mse = 0,1927
- for mixed retweet cascades  $\lambda = 0,01$  with mse = 0,032
- for false retweet cascades  $\lambda = 0,03$  with mse = 0,1931

### REFERENCE

Vosoughi, S., Roy, D., Aral, S.: The spread of true and false news online. Science 359(6380) (2018) 1146–1151