Characterizing Poor Quality Context Specific Content on Online Social Media

Student Name: Karan Grover
Roll Number: 2013048

BTP report submitted in partial fulfillment of the requirements of the Degree of B.Tech. in Computer Science Engineering on 18th April, 2017.

BTP Track: Research

BTP Advisor: Dr. Ponnurangam Kumaraguru

Student's Declaration

I hereby declare that the work presented in the report entitled **"Characterizing Poor Quality Context Specific Content on Online Social Media"** submitted by me for the partial fulfillment of the requirements for the degree of *Bachelor of Technology* in *Computer Science & Engineering* at Indraprastha Institute of Information Technology, Delhi, is an authentic record of my work carried out under guidance of **Dr. Ponnurangam Kumaraguru**. Due acknowledgements have been given in the report to all material used. This work has not been submitted anywhere else for the reward of any other degree.

……. Place & Date: ……...

Karan Grover

Certificate

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

……. Place & Date: …….

Dr. Ponnurangam Kumaraguru

Abstract

In social psychology, a rumor is defined as a story or statement in general circulation without confirmation or certainty to facts. Rumors are known to arise in the context of ambiguity, when the meaning of a situation is not readily apparent or when people feel an acute need for security. Rumors hence are a powerful, pervasive, and persistent force affecting people and groups. Popular online social sites such as Twitter, Facebook and Bebo, have become some of the major news sources as well as the most effective channels for viral marketing nowadays. However, alongside these promising features comes the threat of misinformation propagation which can lead to undesirable effects. This threat is significant because the OSNs have greatly facilitated and accelerated information diffusion processes which makes them a fertile ground for the spread of fake news, misinformation, rumors and hoaxes. The goal of this work is to find characteristics of such rumors on online social media so that they can be possibly identified early in the process.

Keywords: Online Social Networks, Fake News, Rumor, Hoax

Acknowledgements

I would like to most sincerely thank Dr. Ponnurangam Kumaraguru and Prateek Dewan for advising me on this project.

The door to Prof. PK's and Prateek's office was always open whenever I ran into a spot of bother or had a question about the project. They consistently allowed this project to be my own work, but steered me in the right direction whenever they thought I needed it. Their valuable input and guidance helped me define a purpose and well-structured work process together with many valuable discussions.

Introduction

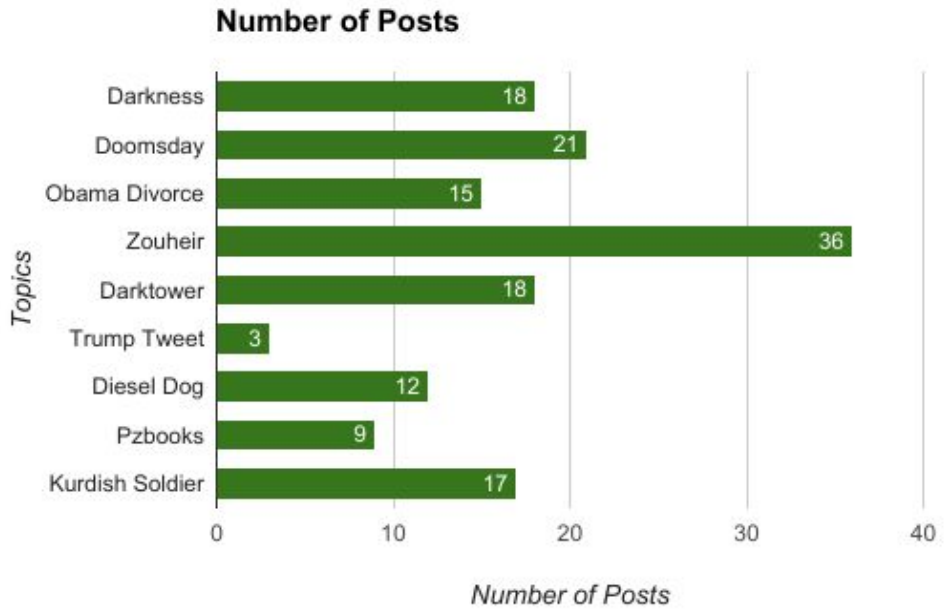
Online social networks provide a rich substrate for rumor propagation. Information received via friends tends to be trusted, and online social networks allow individuals to transmit information to many friends at once. By referencing known rumors from Snopes.com, a popular website documenting memes and urban legends, we track 9 rumors appearing on Facebook. The themes of the rumors vary from Politics and Terrorism to Space and Science. From this sample we infer the rates at which rumors from different categories and of varying truth value are shared and interacted with. We have done temporal and linguistic analysis on the shares and comments and we have measured how the users have reacted to the post over time, how the rumour is debunked over time, how does it affect its probability of being shared again.

Data Collection

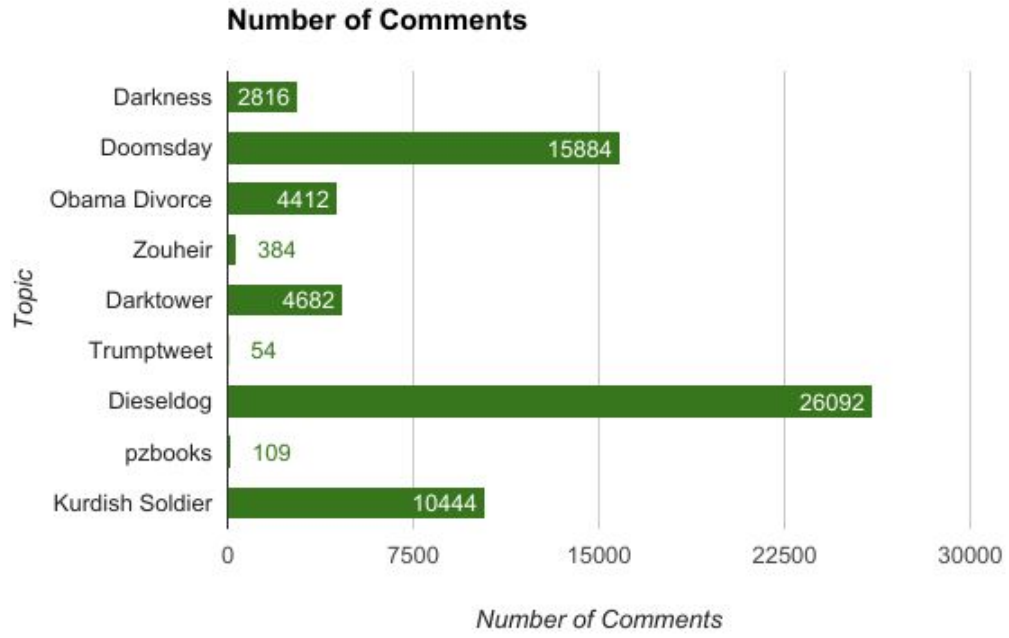
We have total 9 different rumors. They are as follows:

1. November to bring 15 days of Darkness - NASA Claims. The following rumour surfaced in October 2016 and continued to stay on Social Media till around January. It was claimed that an astronomical event between Venus and Jupiter will cause the Sun to dim to a bluish colour, which will take around 14 days to normalize.
2. When Paris turned out it's lights, the rest of the world turned them back on - The following rumor surfaced after the Paris attacks. The rumor was the the lights of the Eiffel Tower were turned off because of the attacks but it was not so. The lights are turned off every night at 1am.
3. Dieseldog - The rumor claimed that a dog named diesel was killed during raids in France when a suicide bomber detonated her explosive vest. The truth is that the dog was killed during a police
4. Doomsday - The post claimed that NASA has confirmed that a Doomsday Asteroid will hit earth on February 16th, but the truth was the it was just a rumor.
5. Kurdish Soldier - The post claimed that a female Kurdish soldier had killed 100 ISIS terrorists. The truth was that the girl in the picture was a journalist who was posing in miltary costume.
6. Obama Divorce - The post claimed that Barack and Michelle Obama had filed for a divorce. But the truth was that it was just a rumor.
7. Pzbooks - A twitter account @PZbooks tweeted about the Paris terrorists attacks before it happened. Contrary to the beliefs that it was intentional, the event was just a coincidence.
8. Trump tweet - This post is a clear example of misinformation. Trump had tweeted a tweet regarding the Charlie Hebdo shooting. However, the tweet resurfaced a few months later and was perceived to be regarding another terrorist attack in Paris. This led to the French Ambassador's response, which could have turned into a political fight.
9. Zouheir - The post claimed that a Muslim guard named Zouheir prevented a suicide bomber to enter into the stadium. However the truth is that he just was just an observer of the event and had nothing to do with it personally.

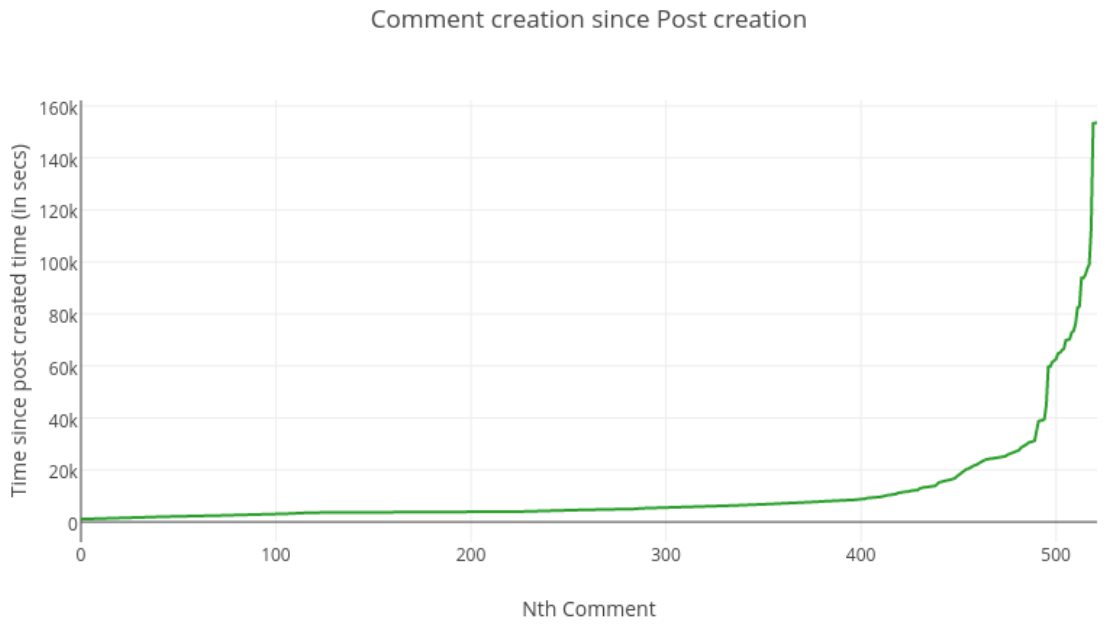
In the above sample we collected a total of 69 posts using the Facebook Graph API. The distribution of posts among different topics is shown below.



To analyze how users have interacted with the fake post, we have analyzed the comments of the post. The total number of comments that we have analyzed are 64877. The distribution of the comments among the post topics is shown below.

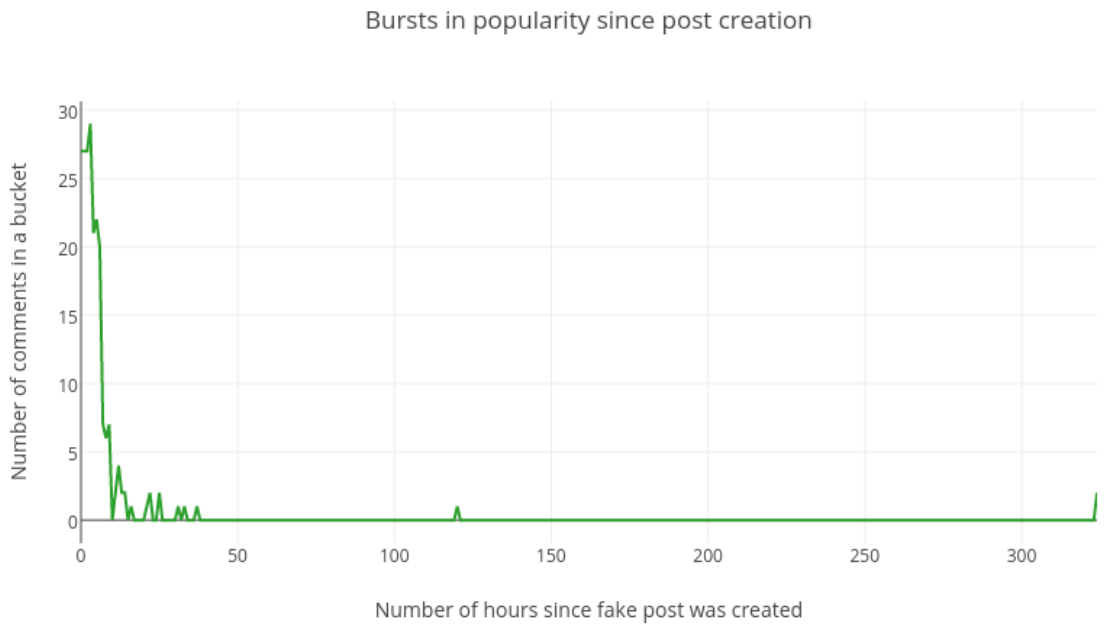


The first analysis that we did was a temporal analysis of the comments. We extracted the timestamps of all the comments and sorted them in ascending order. This is because Facebook's Graph API provides the comments in the order of the time of latest activity and not created time. After sorting the comments by 'created time', we made a plot based on the difference between the time when the comment was posted and when the fake post was posted. Basically measuring how much time after the fake post was posted, was the comment posted. Almost all graphs followed a similar shape as shown below.



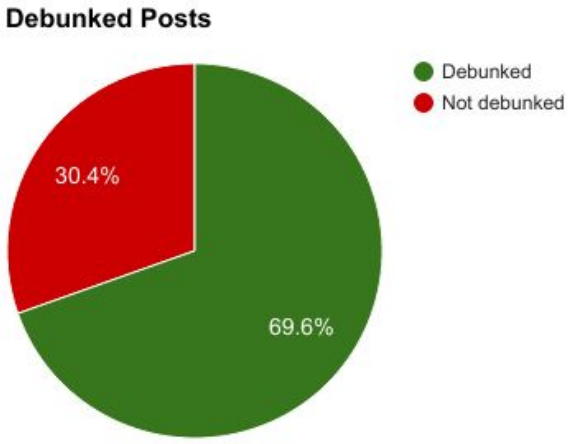
We can see that the above graph looks similar to an exponential curve. It tells us that the interaction with the fake posts happen in short bursts. This means that a lot of people comment and interact with the post in a very short interval of time in the beginning. But gradually the interaction begins to die down as the post started to get debunked or simply due to the passage of time. We can see in the above graph that 90% of the comments are in the first 16% of the time interval and the last 10% of the comments are spread of the latter 84% of the time interval.

To get a more accurate picture we created buckets of 1 hour each since the time the fake post was created and we put the comments into the buckets. We then plotted the graph to visualize the bursts in popularity. The graph for the same post as above, looks as below:

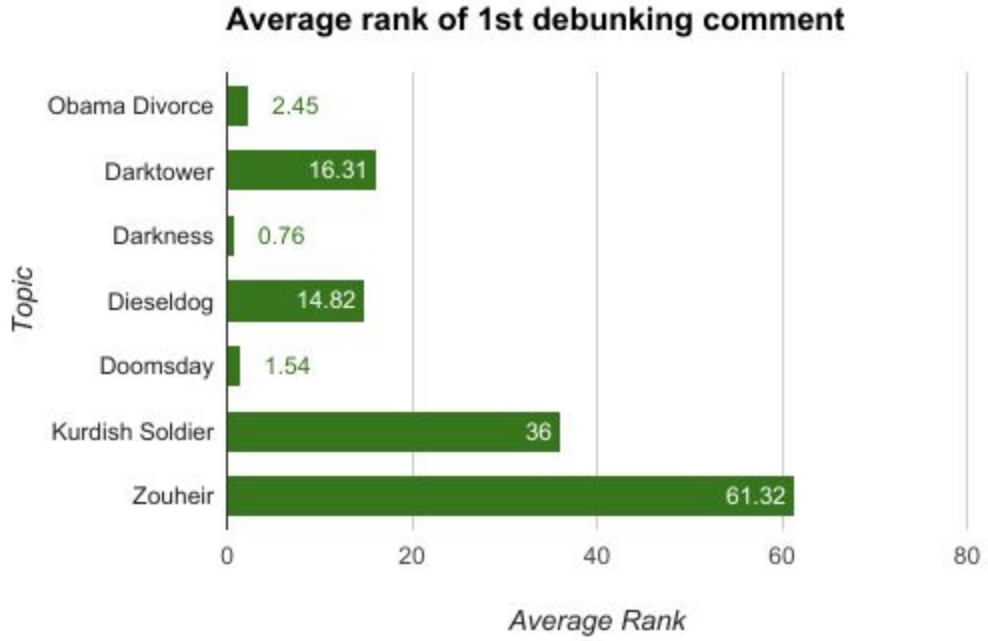


We can see in the above graph that there were 27 comments in the very first hour and drops to around 7 in the 7th hour. We can also check that 98.6% of the comments were posted in the first 40 hours, while the last interaction with the post was 300 hours after it was posted.

The next step was to dive into the text of the comments. The comments can be divided into 2 categories. The first one consists of all the comments the debunk the rumor in the post or question it's legitimacy. The second one consists of all the other comments, most of which are a knee-jerk reaction to the rumor often agreeing with it. We crowdsourced the annotation of all the 64877 comments that we had. There were a total of 1848 comments that debunked the posts. This means that 97.15% of the comments agreed with the posts. If we assume that each comment was by a unique user, this means that 97.15% of the users get duped by such rumors on social media. After the annotation was completed our first analysis was to check what percentage of the posts were debunked in the comments. A post is said to be debunked when there is atleast one comment that claims that the post is fake or questions its legitimacy. We observed that 69.57% of the posts were debunked in the comments. All of the 9 topics had atleast one post which was debunked.

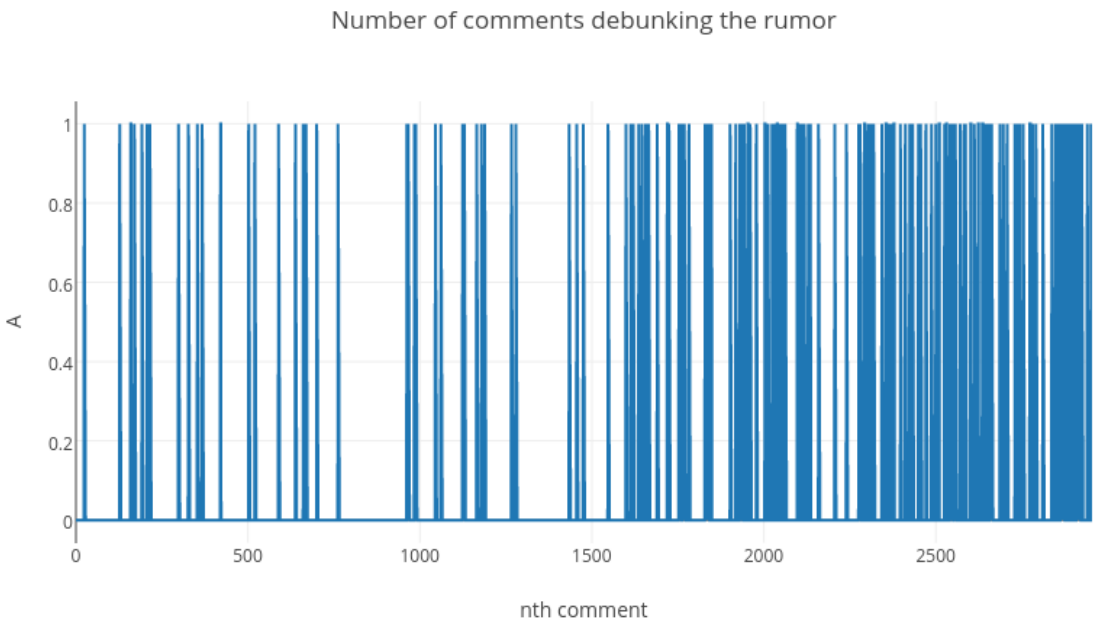


Our next step was to focus on what was the average time period for which these posts existed before a comment claimed that it was fake. Since the different posts are around for different periods of time, it makes more sense to find out the average rank of the 1st comment that claimed the post to be fake. This takes into account the total number of comments, which are a rough equivalent of how long the post have been interacted with.



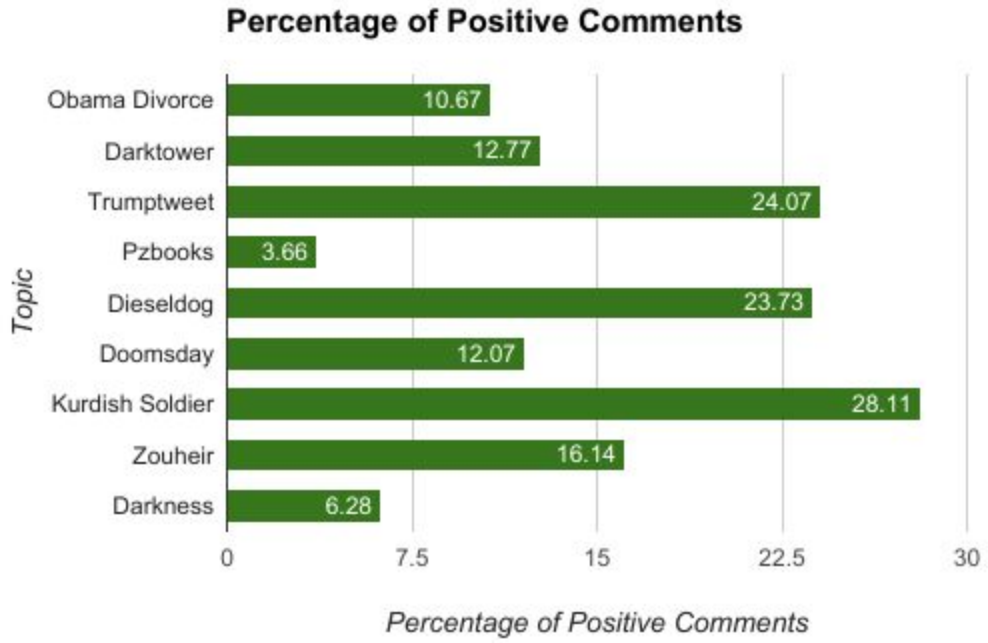
We can see in the above graph that ‘Darkness’ topic was the quickest to be debunked with an average rank of the 1st comment that debunked the post to be 0.76. Clearly a lot of first comments on the post were claiming it to be fake already. This is probably because this rumor tends to come up every year since it’s related to a celestial event. That is why a lot of people already know that this is fake. On the other hand the ‘Zouheir’ post was the one which took the most time to be debunked. This is probably because this was a one-off event and the true event was very similar to the rumor, which confused the people even more.

Our next step was to focus our attention only on the comments that claim the post to be fake. We plotted all the comments of the post against their labels. A label is 1 if the comment claims the post to be fake, and the label is 2 otherwise. The result that we obtained is shown below:

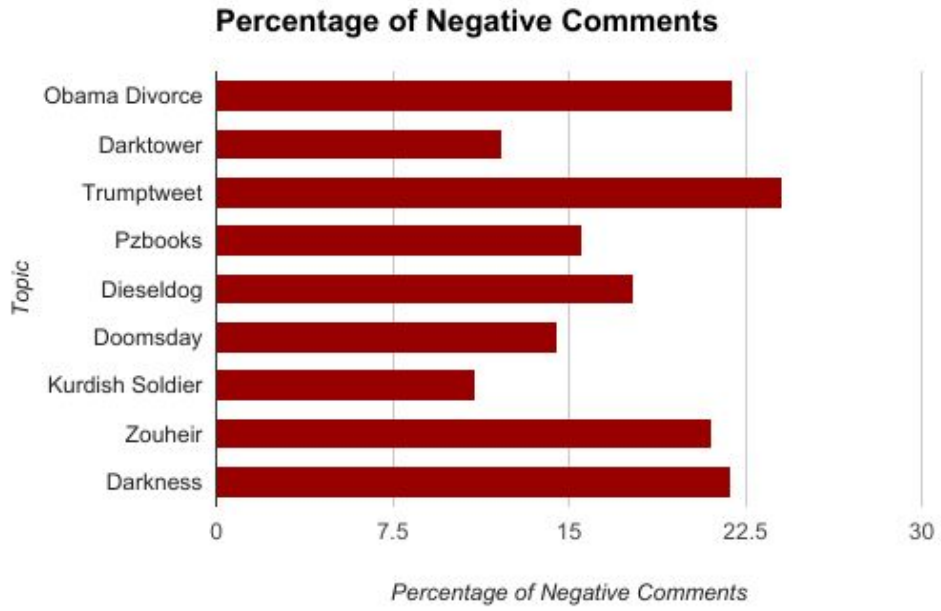


We can see in the above graph that the frequency of comments that claim the news to be fake increases over time. It should be kept in mind that as we noticed above, the number of comments on a post exponentially decreases over time. Hence, the number of comments that claim the post to be fake increase at a very high rate. This is expected for many reasons. First is that, it’s possible as the number of comments that claim the post to be fake increases, more people become aware of its illegitimacy by reading previous comments. This might create a feedback loop in which the more number of comments claiming it to be fake are there, the more people will become aware of it and comment the same. The second reason could be people becoming aware of the rumor from other sources as the rumor gains popularity in online social media and mainstream media.

Our next step was to analyze the sentiment of the comments. Now that we had annotated comments, an even deeper analysis could be done. First we did a normal sentiment analysis of all the comments grouped by topics. The results for the same are shown below.



The graph above shows the percentage of positive comments for each topic.



The above graph shows the percentage of negative comments in every topic. We also observed of all the comments that claimed the post to be fake, 5.61% of the claims are positive and 28.50% of the claims are negative. We also observed that of all the comments in the other category 17.10% of the normal comments are positive and 15.13% of the comments are negative.

Our next step was to analyze the comments that claimed the post to be fake. Below is a word cloud of all the words that are in the comments that claim the post to be fake.

We can see that the words ‘fake’, ‘bullshit’ and ‘news’ come most of the time. We see that around 40% of the comments can be identified using just the keywords: fake, bullshit and hoax.