Comparison between the Twitter Retweet Network and Facebook Like/Comment Network for two Movies: The Hobbit and The Interview

Georgiana Diana Ciocirdel

Polytechnic University of Bucharest, Romania

Abstract. In this paper four different networks are analysed, from two social media portals, Twitter and Facebook. The networks are related to the movies The Hobbit and The Interview. The movie The Hobbit first appeared on the 12/01/2014 in the UK and in the month that followed in the rest of the world. The movie The Interview is quite controversial and has faced many issues before being released solely in the US, first on the 12/11/2014 and then during Christmas time. The analysis has been run on data collected around the release dates. The Twitter networks we are analysing are directed, with edges linking users that have authored a post to users that have retweeted it. The Facebook networks are being analysed in three consecutive steps, marking important release dates around the world of the two movies; these are unimodal, undirected networks, with edges linking users that have liked or commented on the same post made by the official Facebook pages of the two movies. The analysis reveals important things about the four networks: the two Twitter networks differ a lot between them - although big, The Interview's Twitter graph is a "young" one, meaning that authorities and hubs haven't yet been formed properly, the posts are sparse and with very few users retweeting more than one post. In comparison, The Hobbit's Twitter graph has well defined authorities and most of its users have retweeted more than one popular tweet. The Facebook networks also differ: The Hobbit has more users liking and commenting on their posts, especially around the US release, while The Interview network is smaller and doesn't change much in time.

1 Collecting the Data

In order to collect the necessary data for our study, we first tried to use popular data retrieving tools, such as Wolfram Mathematica, NameGenWeb or Social Network Importer for Node XL, but due to some Twitter rate limitations or malfunctioning of the tools, we resumed to writing small python scripts with the use of the Bear Python-Twitter API, an open-source library that provides a "pure Python interface for the Twitter API" and Node XL 1.9.2 for the Facebook queries. Due to the afore mentioned Twitter rate limitations, we were only able to query posts from 12/22/2014 to 12/29/2014 for both movies (as work on this project has been carried out between December 30th and 31st 2014). Tweets

can be either recent, popular or mixed, but we queried only the popular tweets, as it would get us data about more influential users. We then queried for the retweeters of each of the popular tweets. We collected the data in a gdf file format and created a node for each user that has retweeted a popular tweet and a node for each user that authored a popular tweet. We then added an edge between each user who has retweeted a tweet and each user that had authored the specific tweet. The Facebook data was easier to collect and we gathered information as follows:

1. for The Hobbit we collected data from between:
   (a) 12/01/2014-12/02/2014, which marked the release dates in the UK and France for the London and Paris premieres;
   (b) 12/04/2014-12/09/2014, which marked the release dates in the USA for the New York and Los Angeles premieres;
2. for The Interview we collected data from between:
   (a) 12/11/2014, which marked the Los Angeles premiere in the USA;
   (b) 12/24/2014, which marked the internet release in the USA;
   (c) 12/25/2014-12/26/2014, which marked the release date in the rest of the USA.

In the resulting graphml files, we have created a node for each user that has liked a post or commented on a post on the specific Facebook page. We then added an edge between each user who has liked/commented on posts on the page.

2 Facebook and Twitter Network Analysis

In this part we are going to analyse the four networks formed with the data we have retrieved as presented in the previous part.

2.1 The Network Structure

The two Facebook networks were taken from the official Facebook pages of the two movies, The Hobbit and The Interview. Each node represents a user and the links between two users are formed when both have liked or commented on the same post made by the page. Therefore, this is an undirected network.

The two Twitter networks were formed from the search results using the queries "the hobbit" and "the interview". We have considered the most popular posts, i.e. the "trending topics". We have added a node to the network for each user that has either posted or retweeted a post and then added a directed edge from the user that has authored the tweet (the source) to the user that has retweeted it (the target). Therefore, this is a directed network.

In this section we will analyse the four networks by interpreting the visual layouts (the graph representation) and by combining certain metrics so as to obtain a mathematical insight into the data. For the visual representation of the networks, we have used Gephi.

2.2 The Interview - the Twitter Network

Raw Results

Metric	*Value*
Network Type	Directed
Number of Nodes	7 872
Number of Edges	8 152
Average Degree	1.036
Network Density	0
Diameter	2
Connected Components	13

Betweenness Centrality We have used the Fruchterman-Reingold algorithm in order to produce a nice visual network of all the data we have gathered from Twitter. Then, we have enlarged the nodes with the highest betweenness centrality in the directed network (1). This measures how often a node appears on shortest paths between other nodes in the network, in other words this highlights the Twitter users that have the greatest number of retweets or the Twitter users that get their tweets retweeted in cascade. We can easily notice how certain areas are denser than others - these are popular tweets that have been retweeted by many other users, so the central node points to many other nodes. There are many people who have retweeted one popular tweet, but there aren't many people who have retweeted many popular tweets.

Unsurprisingly, the nodes with the greatest betweenness centrality correspond to users like Seth Rogen and James Franco (who play the main characters in the movie), The Interview (the official Twitter user of the movie), Hollywood Reporter (an American media publication) or Variety, (an American magazine).

Communities in the Network A community on Twitter could be defined as a set of users that have more links within the set than outside of it, in other words, users that share quite similar views or prefer similar tweets and therefore, retweet them. Figure 2 shows the network colored based on different communities and the users with the highest numbers of followers have been enlarged. These results have been yielded when determining communities with a 20.0 modularity (we wanted to highlight only the most important communities). We can see from these layouts that the communities are rather fuzzy and intertwined, but a very important thing that can be deducted is the fact that one or more users with a huge number of followers (influential users) are part of each of the biggest communities highlighted (the pink, the blue or the green community). Moreover, we can now clearly see how the ideas are spread from one small group to the other: there are users that retweet one tweet in their small group; one of the retweeters also gets retweeted by one other user, which then gets retweeted by many other users in his/her small group. This is, in fact, the way the Twitter network functions.

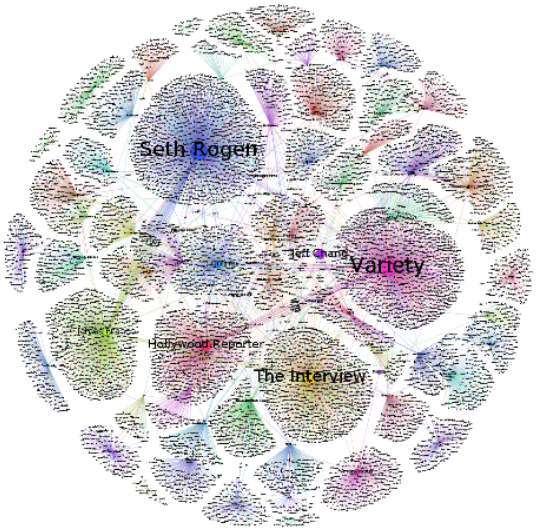

Fig. 1. Highest betweenness centrality in the Fruchterman-Reingold layout for the Twitter network of The Interview

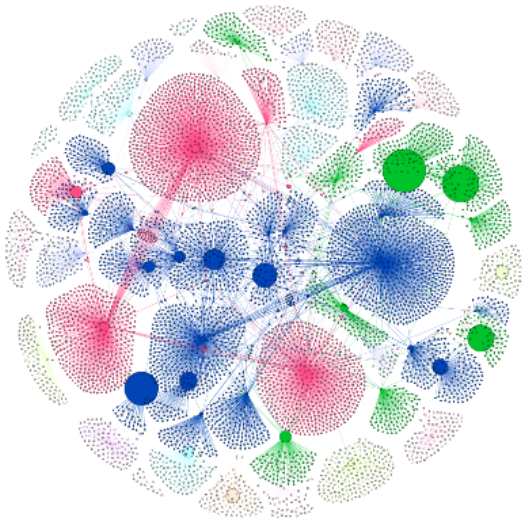

Fig. 2. Communities in the Fruchterman-Reingold layout for the Twitter network of The Interview

2.3 The Interview - the Facebook Network

Raw Results

Metric	*Network*	*Value*
Network Type	12/11-12/2014	Undirected
	12/24/2014	Undirected
	12/25-26/2014	Undirected
Number of Nodes	12/11-12/2014	301
	12/24/2014	263
	12/25-26/2014	388
Number of Edges	12/11-12/2014	7 742
	12/24/2014	6 689
	12/25-26/2014	10 609
Average Degree	12/11-12/2014	51.442
	12/24/2014	50.87
	12/25-26/2014	54.686
Network Density	12/11-12/2014	0.171
	12/24/2014	0.194
	12/25-26/2014	0.141
Diameter	12/11-12/2014	4
	12/24/2014	2
	12/25-26/2014	4
Connected Components	12/11-12/2014	1
	12/24/2014	3
	12/25-26/2014	2

Betweenness Centrality We have used the Fruchterman-Reingold algorithm in order to produce a nice visual network of all the data we have gathered from Facebook. Then, we have enlarged the nodes with the highest betweenness centrality in the undirected network (3). For the Facebook network this shows the users that have liked or commented on more than one post in the same day, so users who are brokers between two or more clusters in the network.

One can easily notice that there is no major difference between the three instances of the network. In all cases, there are users who have liked more than one post and users who have liked only one, so the bigger nodes represent brokers between the communities that have formed around each post.

Communities in the Network A community on the Facebook page of the movie could be defined as a set of users that have liked or shared one or more similar posts. Figure 4 shows the network colored based on different communities and the users with the highest betweenness centrality have been enlarged. These results have been yielded when determining communities with a 1.0 modularity. We can see from these layouts that the communities are very well defined and very well separated from one another. It would make sense to assume that

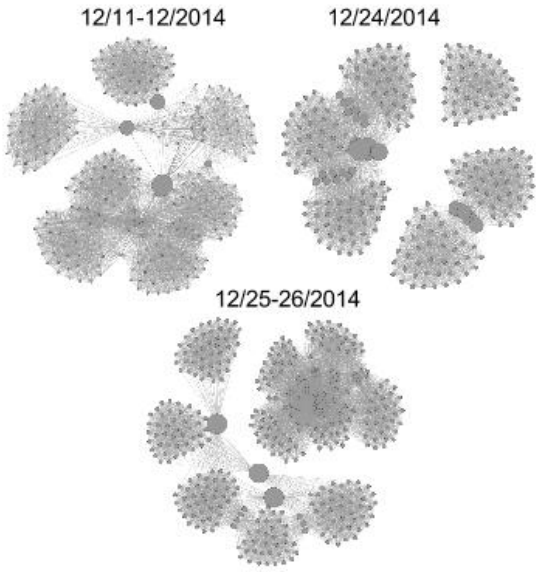

Fig. 3. Highest betweenness centrality in the Fruchterman-Reingold layout for the Facebook network of The Interview

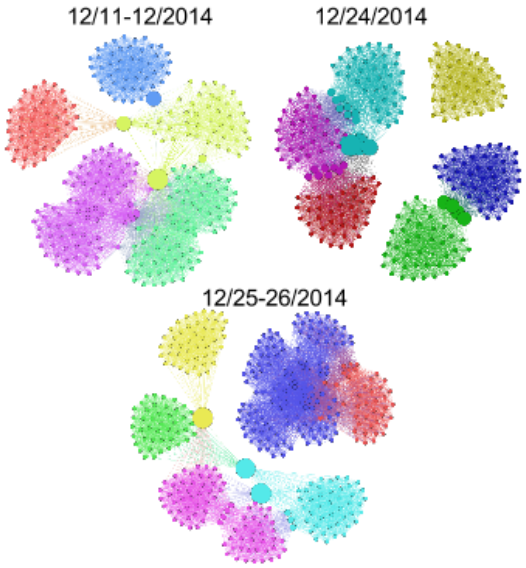

Fig. 4. Communities in the Fruchterman-Reingold layout for the Facebook network of The Interview

one user who has liked the Facebook page of the movie and has then liked or commented on a post, has maybe received the post on their timeline or has browsed through the page's previous posts. It would also make sense to assume that someone who has liked one post might have liked at least another one, but if we look at the diagrams, it is clear that there aren't many users who have done this (there aren't many brokers in the network). This could be the consequence of the manner in which the Facebook timeline shows the posts (it doesn't often show two updates of the same page in a row) or simply because one page cannot publish that many important notes per day (or during a two-day timeline).

Connected Components The connected components are displayed in figure 5. The network obtained between the 11th and the 12th presents only one connected component. This might suggest that the page's posts were all very interesting and/or important, so there were people who have liked or commented on more than one.

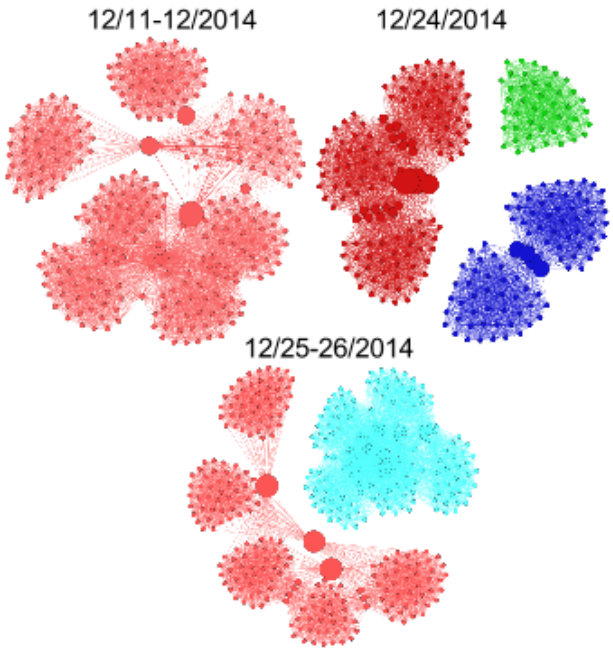

Fig. 5. Connected components in the Fruchterman-Reingold layout for the Facebook network of The Interview

2.4 The Hobbit - the Twitter Network

Raw Results

Metric	*Value*
Network Type	Directed
Number of Nodes	2 646
Number of Edges	2 908
Average Degree	1.099
Network Density	0
Diameter	2
Connected Components	5

Betweenness Centrality We have again used the Fruchterman-Reingold Algorithm in order to produce a nice visual network of all the data we have gathered from Twitter. Figure 6 shows the Fruchterman-Reingold layout of the directed Twitter network. The nodes with the highest betweenness centrality have been enlarged. This Twitter network has a similar visual structure as the previous one. We can easily see on the margin of the circle users whose posts get retweeted only once. However, in the center of the circle, there is quite a dense network of users that got their tweets shared multiple times in a row. The nodes with

Fig. 6. Highest betweenness centrality in the Fruchterman-Reingold layout for the Twitter network of The Hobbit

the greatest betweenness centrality correspond to users like The Hobbit Movies UK, The Hobbit, Major Cineplex (an operator of movie theaters in Thailand) again the Variety magazine and a user named Marlin Stolsig, which we could not identify. Unlike the Twitter network of 'The Interview', the most popular users in this network are spread across the world and are not only localized in the US.

Communities in the Network Figure 7 shows the network colored based on different communities and the users with the highest numbers of followers have been enlarged. These results have been yielded when determining communities with a 20.0 modularity (we wanted to highlight only the most important communities).

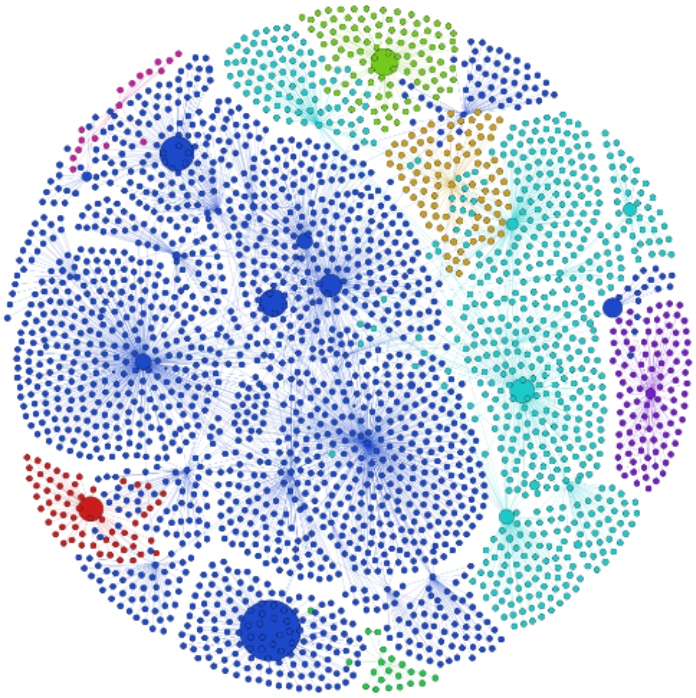

Fig. 7. Communities in the Fruchterman-Reingold layout for the Twitter network of The Hobbit

2.5 The Hobbit - the Facebook Network

Raw Results

Metric	*Network*	*Value*
Network Type	12/01-02/2014	Undirected
	12/04-09/2014	Undirected
	12/10-12/2014	Undirected
Number of Nodes	12/01-02/2014	468
	12/04-09/2014	854
	12/10-12/2014	290
Number of Edges	12/01-02/2014	12 299
	12/04-09/2014	23 251
	12/10-12/2014	7 110
Average Degree	12/01-02/2014	52.56
	12/04-09/2014	54.452
	12/10-12/2014	49.034
Network Density	12/01-02/2014	0.113
	12/04-09/2014	0.046
	12/10-12/2014	0.17
Diameter	12/01-02/2014	3
	12/04-09/2014	5
	12/10-12/2014	2
Connected Components	12/01-02/2014	5
	12/04-09/2014	2
	12/10-12/2014	3

Betweenness Centrality We have used the Fruchterman-Reingold Algorithm so as to produce a visual representation of the network in time. Figure 8 shows the Fruchterman-Reingold layout of the undirected Facebook network. As both the metrics and the visualizations suggest, the three instances of the network are different from one another: we can see that the first two networks are more connected than the last one, meaning that in the first two networks (and more in the second than in the first), there are many users who have liked or commented on more than one post, so the clusters that have formed around each post are linked to one another. Let us not forget that the movie has premiered in London and Paris on the 1st and 2nd of December, then during the 4th and 9th it has premiered in the US and between the 10th and 12th it has been released in other parts of Europe. The three networks built around these dates might suggest that The Hobbit was more popular in the US than in Europe. We can also notice that the "brokers" in the second network are smaller in size, which means that less shortest paths pass through them than in the other networks. This happens because there are more central nodes in the second network than in the other two, so "the traffic" is split among them. Again, this suggests that the Facebook page of the movie has become more popular during the release dates in the US than in Europe.

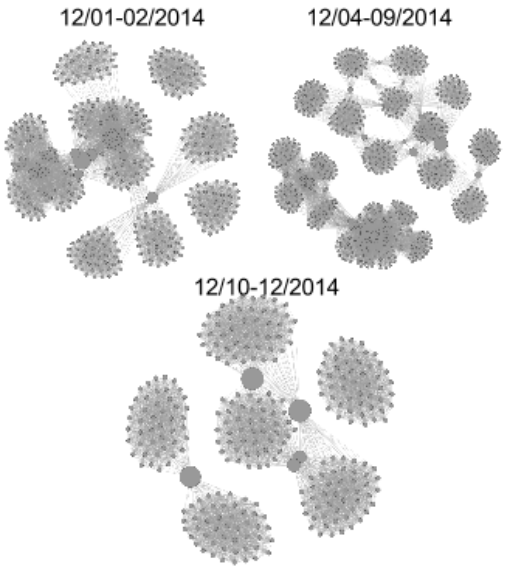

Fig. 8. Highest betweenness centrality in the Fruchterman-Reingold layout for the Facebook network of The Hobbit

Communities in the Network Figure 9 shows the network colored based on different communities and the users with the highest highest betweenness centrality have been enlarged. These results have been yielded when determining communities with a 1.0 modularity. Just like the networks for the movie The Interview, the communities here are very well defined and very well separated from one another. This time, however we can see some differences between the three networks: for the first and second graph, the communities have engulfed more than one post (which again proves how popular the page was during this time frame); the third network, however is smaller and the communities are only defined around single posts, with a few brokers between them.

Connected Components The connected components are displayed in figure 10. We can see that the first and the third network contain five, respectively three connected components, whereas the second one contains only two.

3 Conclusions

The analysis of the four networks was based on data collected from the Facebook and Twitter social web pages. Using free services and collecting the data between the previously mentioned dates has lead to less relevant information than we would have desired. Therefore, the graphs we have obtained are smaller than in

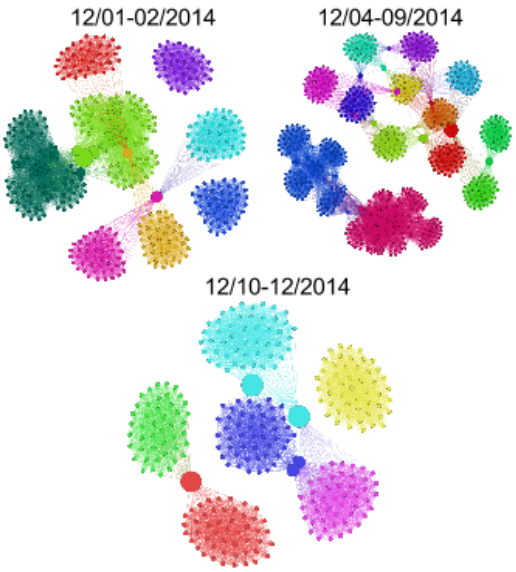

Fig. 9. Communities in the Fruchterman-Reingold layout for the Facebook network of The Hobbit

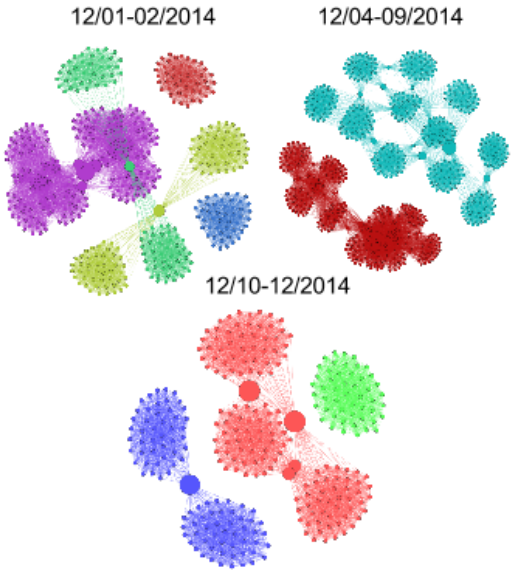

Fig. 10. Connected components in the Fruchterman-Reingold layout for the Facebook network of The Hobbit

reality. We have improved the quality of the data, by querying the most popular tweets on Twitter and retrieving data from the release days on Facebook.

There are several differences between the networks we have analysed and we have highlighted these in the previous paragraphs. The most important would be:

1. network structure - While the Twitter networks are directed, the Facebook ones are undirected. In the Twitter network, we have linked users to other users whose posts they have retweeted, whereas in the Facebook network there is an edge between all the users that were active around a post.
2. data range - We have analysed the Facebook network during three different periods of time and with different time span for the two movies. Data for the Twitter networks has been gathered over the time span of one week and is not directly related to the release dates of the movies (this is due to the rate limitations).
3. network analysis - The Interview Twitter network seems "younger" than The Hobbit's: the hubs and the authorities are not yet well defined and the graph is rather sparse - there aren't many users who have retweeted more than one post. Regarding the Facebook networks of the two movies, The Hobbit has more users liking and commenting on their posts, especially around the US release, while The Interview network is smaller and doesn't change much in time.