Exploring Tweets in the Aftermath of Typhoon Haiyan

Introduction

Typhoon Haiyan was a tropical cyclone that originated in the Pacific Ocean and tracked into the South China Sea. On November 6, 2013, the storm hit the Philippine Islands, causing widespread devastation. The storm brought extremely high sustained winds (estimated at 145 mph) and large waves (excess of 15 ft). Although the extent of the damages are still being assessed, it is clear that the impact of this disaster is extreme. The death toll is estimated to be over 2,000 with over 800,000 displaced refugees.

Data

Social networks have changed the way that information is exchanged. In the case of disasters such as Typhoon Haiyan, users from all demographics and locations can express their reactions. To explore this phenomenon, I performed a Twitter search on the term “Typhoon Haiyan” through the Social Network Importer extension for NodeXL. The tweets range from 16:07 to 16:27 UTC on November 9, 2013. An overview of the data can be seen in Figure 1 below. The vertices are sized by betweenness centrality and colored by in-degree. Twitter data is interesting in that Twitter encourages immediate, short messages. Tweets are often utilized to express reactions, rather than reflections or researched news. In addition, tweets reach large audiences through retweets, in which users copy a message so that it displays to their followers. This often makes Twitter activity in response to an important event a point of social media and information flow analysis.
Headline 1: Influence from the Pope and the Astronaut

Manipulation of the visualization shows that there are two people that largely influenced the Twitter communication: Pope Francis and NASA astronaut Karen Nyberg. Pope Francis tweeted a request for prayer for the victims of Typhoon Haiyan. I captured 269 retweets during this capture process. A viewing of the tweet on Twitter indicates 20,540 official retweets as of this writing. Karen Nyberg is a NASA astronaut aboard the International Space Station. She tweeted an image of Typhoon Haiyan as viewed from space. I captured 178 retweets through the NodeXL importer. Twitter shows 4,148 official retweets of the image. It should be mentioned that both of these influence tweets were also unofficially retweeted a large number of times.

Figure 2 is a visualization meant to demonstrate the influence of Pope Francis and Karen Nyberg. The visualization has been filtered to remove vertices with less than 1 in-degree. This is meant to remove
twitter handles that did not solicit any replies or retweets in order to clear clutter from the visualization. The vertices were then sized according to in-degree to emphasize twitter handles whose tweets received many replies or retweets. Finally, the vertices are colored by betweenness centrality.

Figure 2: Nodes with High In-Degree

The size of the vertices clearly shows the influence Pope Francis and Karen Nyberg. These vertices are much larger than any of the other nodes. Other notable vertices include Yoko Ono, the Philippine Red Cross, and BBC World.

Headline 2: Retweeting Groups

I observed the data set using the Clauset-Newman-Moore grouping algorithm included in NodeXL. I also used the autofill to size the vertices by the in-degree. The goal was to observe clusters of nodes. My hypothesis was that each group would have one or two primary Twitter handles with many retweeters. The results can be seen in Figure 3.

Created with NodeXL (http://nodexl.codeplex.com)
I grouped all of the standalone tweets into a single group and collapsed it into the bottom right. This removes some noise and makes the other groups more clear. We can see that my hypothesis seems to be accurate. The largest group consists almost entirely of retweets from Pope Francis’ one tweet asking for prayers. The next largest group consists mostly of retweets of Karen Nyberg’s image of the typhoon from space. However, it two fans to the left and right from BuzzFeed News and NBC News. We see this common theme of one tweet with many retweeters and a few fans to the side in many of the groups.
Figure 4 shows a grouping with the same specifications as above, but with edges combined. It is clearer through this visualization that the groups are mostly isolated. The few links between groups are relatively small.

Headline 3: Those Close or Affected by the Event are More Concerned with Impact than Appearances

I next looked into the activity of Twitter users near the Philippines, the most populace place the typhoon had impacted. I filtered the Twitter handles according to time zone, fading the other vertices and edges
to only 10% opacity. The results can be seen in Figure 5.

I found that the most influential Twitter handle from this time zone is philredcross, which is the Phillippine Red Cross. Other notable vertices include Philippine singer and actress KC Concepcion, who is also a United Nation’s World Food Program ambassador, as well as Rappler, a Phillippine based social news network. The reason behind altering the opacity of the filtered data instead of hiding completely is that we can still identify edges between vertices in the Philippine time zone to those outside of this time zone. We can see a large cluster of vertices to the lower right of the center, indicating that many of these people retweeted Pope Francis’ call for prayer. In addition, we can see a small cluster in the upper left retweeting a message from Korean actor Minho Lee. The filtered visualization with nodes separated into groups can be seen in Figure 6 below.

This visualization was meant to get an idea of the activity from those affected by the storm compared to those further away. In Figure 6, we identify the G2 group as the retweeting group for Pope Francis’
tweet and G3 as the group retweeting Karen Nyberg’s image. We can see that the tweet from Pope Francis is influential, just as it is for the larger data set. We can also see that the tweet from Karen Nyberg has almost few retweets from this time zone. I hypothesize that the people closer to the storm’s impact area are more focused on the emotional impact and methods to help than by the superficial or headline focused tweets.

Figure 6: Groups Filtered to Philippine Time Zone
Critique of NodeXL

NodeXL proved to be a very involved and detailed tool. It has a number of different functions that make it very useful in visualizing a Twitter network. The Social Media Importer add on proved easy to use, however I did find it frustrating that I was not able to specify a time based search range for the Twitter search. I would have liked to search for tweets at different points during the event, but I do not see a way to achieve this if I were not collecting data live. Displaying of groups and edges proved to be very useful and intuitive. However, I would have liked to have easier control of the coloring options.

There were several options I would have liked to have. I wanted to label each group by the top mentioned, but was only able to label with the top 10 mentioned, which ended up being crowded and not very useful. It was also difficult to select groups and collapse them. A useful feature would have been the ability to select all groups with fewer than X number of edges or vertices in order to collapse them. The fact that groups will not stay collapsed when performing certain operations emphasized this problem.