The Walking Dead, Hershel, and Twitter: An analysis of tweets surrounding a featured character trending in popular media.
Matthew Mauriello – mattm@cs.umd.edu – 11/12/2013

Introduction

The world is experiencing an epidemic obsession with zombie culture that even researchers are having fun with [2]. This project seeks to look at a small sub-culture that has formed around the television depiction of famed graphic novel series “The Walking Dead”, which has captured the attention of the masses each Sunday evening. This zombie fueled franchise’s recent episode (i.e. airdate: 11/10/2013) focused primarily on the character of “Hershel”, which is unique for this series as he is part of the supporting cast. The increased attention on this character caused the term “Hershel” to begin trending on Twitter and it was at this time that tweets were captured using NodeXL. The analysis in this project focuses on the structure of the network created by tweets mentioning “Hershel”, the users that are part of this conversation, and finally the word-pairs that shed light on their sentiments.

Dataset

The dataset was created using the “import Twitter Search Network” option of NodeXL, which created an undirected network of 809 vertices connected by 734 edges. The network is sparse with a graph density of ~0.0016, but the average geodesic distance between vertices is 3.59 due to the nature of the “Hershel” conversation.

Overview:

An overview of the data indicates that the network of “Hershel” tweets surrounds two of the vertices, which are unsurprisingly the “walkingdead_amc” and “amctalkingdead”. For those not in the know, these two vertices are the accounts of the show itself and the late night talk show that discusses each episode. We can readily see a large fan base communicating with these two shows via Twitter. However, the users creating this network remain a mystery and what we would really like to know about is who bridges the two primary vertices, where in the world are these users tweeting from, and what sentiments are they expressing about “Hershel”.

Figure 1: Overview of “Hershel” trending tweet network.
Headline 1 – There are few bridges between the two primary conversational vertices.

Figure 2 was created by setting vertex label and shape to the twitter user’s name and picture respectively. The size of the vertices was scaled from 1.5 to 500.0 based on the calculated betweenness centrality. Dynamic filtering (betweenness centrality >= 2.0) was then applied to locate the network bridges.

The overview in Figure 1 represents all conversations surrounding the character of “Hershel” during the broadcast of this week’s episode and the follow-up talk show. We can easily identify that the “walkingdead_amc” and “amctalkingdead” vertices serve as the primary bridges between all nodes in this network. However, the dynamic scaling shows that the fan base is highly polarized between the two different incarnations of the franchise’s television fandom. Additionally, we see that there are only two vertices that connect the two primary vertices (beth_seed and beckjospray). Now that the graph has been filtered, we can also see that there are three types of conversations contributing to the trending term: single users mentioning the search term, users mentioning and including the primary bridge
vertices in their tweets, and entirely separate networks conversing about the search term but not including the primary bridge vertices.

**Headline 2 – The conversation is global.**

Though the show is airing during prime time television hours in the United States, we can see that the airing of this is truly a global event when grouping the users by the self-report data contained within the collected tweet attributes. The location attribute pulled from Twitter users is often not reliable because some users like to set their location to nonsensical locations (i.e. “Wonderland”, “Behind you”, and “Fly-over Country”). Additionally, Groups are difficult to form using the location attribute because there are as many locations as there are vertices. A more reliable and easier to group by attribute is the user’s time zone, which organizes the users into 35 distinct geographic areas and 1 unknown group shown in Figure 3.

![Figure 3: Time zones of the “Hershel” trending tweet network.](image)

Figure 3 shows large interconnected clusters across the continental United States. However, we can see a there is a global population highlighted by the red box. Many of these groups are self-contained networks discussing the show from locations around the world that includes: Brasilia, Central America, Santiago, Berlin, Amsterdam, Caracas, Greenland, Casablanca, Newfoundland, Rome, and Tokyo. Besides
the fact that this conversation is global, another interesting point highlighted by this grouping is the small network in Quito, Ecuador talking about “Hershel” with the primary vertices and among themselves; this appears to be the largest cluster outside of the continental United States that we are able to detect.

![Figure 4: Time zones of the “Hershel” trending tweet network with edges combined.](image)

In Figure 3 it is a bit difficult to see what groups are talking to other groups. By turning on the combine edges setting in NodeXL’s layout options, we can see that much of the conversation occurs within the continental United States and the unknown categories. The global conversation contributors primarily talk amongst themselves. It is possible that these conversations are occurring this way due to language barriers as a manual examination of tweets from these clusters shows regional dialect and languages being used. For example, one user writes:

“RT @p4d4ckles: Resumindo o ep: HERSHEL @#IS PRISAO DETONADA, ZUMBIS, GOVERNADOR!”  **Translation:**

RT @ p4d4ckles: Summarizing the episode: Hershel <explicative>, Prison Detonated, Zombies, Governor!
However, it is likely that some (or all) of these global fans speak and understand English by virtue of watching the show when it is being originally aired – this suggests that some of these nodes are acting as cultural bridges and adding global diversity [1] to the “Hershel” conversation.

**Headline 3 – Clusters have slightly different conversational sentiments about the episode.**

Figure 4 is example of clustering the tweets using the Clauset-Newman-Moore algorithm and Fruchterman-Reingo layout. Vertices without neighbors (the big blue circle in the top left quadrant) have been compressed into a single group and then all but the three largest clusters have been collapsed.

It is from these clusters that we begin to look at word-pairs to get a sense of what these groups are talking about. Mohtarami *et al* [3] suggests that word-pairs can be used to perform sentiment analysis when rigorous statistical models, originally generated from manual annotations and tags, can be applied to relate pairs to emotional state predictions. The analysis in this project is not as rigorous but modeled after Mohtarami’s original manual tagging. Table 1 below shows the word-pairs for each cluster and was generated by the graph metrics feature of NodeXL.
<table>
<thead>
<tr>
<th>Word-Pairs G2</th>
<th>Count</th>
<th>Word-Pairs G3</th>
<th>Count</th>
<th>Word-Pairs G4</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>rt,walkingdead_amc</td>
<td>89</td>
<td>amctalkingdead,hershel</td>
<td>25</td>
<td>rt,immortaldixon</td>
<td>6</td>
</tr>
<tr>
<td>hell,yeah</td>
<td>86</td>
<td>amctalkingdead,think</td>
<td>21</td>
<td>immortaldixon, hershel</td>
<td>6</td>
</tr>
<tr>
<td>walkingdead_amc,hell</td>
<td>85</td>
<td>think,hershel</td>
<td>14</td>
<td>hershel,scott</td>
<td>6</td>
</tr>
<tr>
<td>yeah,hershel</td>
<td>85</td>
<td>gotten,sick</td>
<td>7</td>
<td>scott,sir</td>
<td>6</td>
</tr>
<tr>
<td>walkingdead_amc,stay</td>
<td>5</td>
<td>hershel,killed</td>
<td>5</td>
<td>sir,officially,badass</td>
<td>6</td>
</tr>
<tr>
<td>stay,help</td>
<td>5</td>
<td>hershel,immune</td>
<td>5</td>
<td>officially,badass,regards</td>
<td>6</td>
</tr>
<tr>
<td>help,hershel</td>
<td>5</td>
<td>thewalkingdead,hershel</td>
<td>5</td>
<td>badass,regards</td>
<td>6</td>
</tr>
<tr>
<td>hershel,thewalkingdead</td>
<td>5</td>
<td>turnandcough, thewalkingdead</td>
<td>4</td>
<td>regards,entire</td>
<td>6</td>
</tr>
<tr>
<td>thewalkingdead, walkingdead_amc</td>
<td>4</td>
<td>hershel,carl</td>
<td>4</td>
<td>entire,twd</td>
<td>6</td>
</tr>
<tr>
<td>bad,ass</td>
<td>4</td>
<td>carl,risk</td>
<td>4</td>
<td>twd,fandom</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 1: Word-pairs and counts generated from the expanded clusters in Figure 5.

A manual analysis of Table 1 word-pairs would indicate that the three largest clusters (G2, G3, and G4) all express slightly different emotional responses. G3, centered on the post-show “amctalkingdead”, discusses the results of “Hershel’s” actions and condition during the episode while also expressing concern about “Carl” - another controversial character within the series. This reaction can be contrasted by users in G2 who are anchored more to “walkingdead_amc” and focus more on “Hershel” doing the right thing by helping those in need as well as the characters viral quote “hell yeah” as illustrated by G2’s most popular URL: https://vine.co/v/hTKOr3wtVQ. An analysis of G4, somewhat limited by its small population, indicates communications to a possible local bridge (i.e. the “immortaldixon” user account) and that the network primarily considers “Hershel” to be a “bad ass”. To use Mohtarami’s affective emotional states, G2 would be tagged as feeling a “cherished” emotion for their sentiments about “Hershel’s” actions; G3 and G4 would likely be tagged with experiencing “excitement” over the final results of “Hershel’s” actions in the episode.

NodeXL Critique

Using NodeXL to perform network analysis is excellent. Importing data from Twitter is very easy; however, the download of tweets is throttled down (by Twitter), it generally pauses, and the resulting error messages after the pause are rather confusing. The fact that NodeXL is integrated with Excel and that many of the processes that users would apply post-import are automated is perhaps one of the best features. The interface to NodeXL is visual, which makes it easy for a wide range of users to use it and create impactful visualizations. For this particular project, running NodeXL on a Windows 7 machine using Excel 2010 was very stable and did not exhibit any of the frustrations reported by other users.

Additional features might include an interface with more onboarding because learning NodeXL involved a lot of reading as well as trial and error. Additionally, this onboarding interface could include suggestions about how to analyze the loaded dataset for novice users. There is also some conflict between settings in the graph and clustering options being overwritten by what is contained within Excel, although this is easy to correct once the user realizes what is happening – they do take some
digging to find; a conflict indicator somewhere in the interface would be helpful as well as automating the resolve process. Finally, it is really difficult to undo changes and this resulted in constantly re-running the same series of actions; though this could possibly be handled by the macro feature of Excel. Ultimately, the experience of using NodeXL was very good and the resources available for learning the tool are very extensive for those willing to search, read, and play with the software.

References

