POLITICAL PREFERENCES OF MASS MEDIA

As an international student, it is fun to witness the whole election process. With
President Obama being re-elected, and with his photo of hugging the first lady being the
top retweeted tweet, I start to wonder how politically biased are the mass media all
around us in this country?

DATA

First, I hunted for a complete list of media in the United States. At www.usnpl.com, I
found a (almost) complete list of newspapers press, along with a list of TV stations, with
twitter IDs, by states, in the United States. There are 2287 newspaper presses and 917
TV stations in the lists.

After several attempts of getting a complete network of these presses, I realized that it is
intractable to try to import such a huge network from twitter using the importing tool
given the current Twitter rate limits.

I then switched my strategy and focused on a small subset of these media. In a previous
study by Prof. Jen Golbeck, twenty media outlets in Twitter were evaluated and an
average score on political preference of the audiences of these media outlets were given.

That score could be used to estimate the political preferences of mass media based on
their following/followed relationships by propagating and getting the mean of the
scores. Those that has no connection to the top 10 news outlets are discarded.

Finally, with the help of Professor Shneiderman, I successfully downloaded the data set,
which contains all the connections among all the twitter accounts of mass media in this
country. There's originally 2174 vertices in the graph, and 16878 edges among them.
After removing irrelevant edges and vertices, there are 1663 vertices and 15301 edges
left.
First I tried a spiral layout. In order to make it more visible and easy to interpret, I lowered the opacity of edges to 20%. The vertex colors are from the grouping of states. Each edge represents a following relationship.

It is observed that the picked ten top news outlets are all around the center of the graph. Across the whole spiral, no obvious differences in influences of media by states are observed.
Then I tried a different layout and color set. In this graph I used circle layout and I manually put the ten top news outlets in the middle. The color is auto-filled by the normalized predicted P-scores. The predicted P-scores are calculated in 4 passes based on the available P-scores of neighbor vertices.

It seems that even though the edges are denser in the middle, that doesn’t represent all the patterns in this graph, as the expected radial pattern doesn’t dominate other patterns. Apparently there are other patterns in the graphs. I suppose that could be due to media twitter accounts in the same state scattered all around the graph and the high locality in the connections. Similar patterns could be observed in the following graph (grouped by states) as well.
In this graph, I grouped all the vertices by states and used their P scores to color the nodes. Connections are further dimmed in order to make it easy to observe.

It seems that those solid states, such as the "blue west coast" states, are still mostly blue, and it also applies to solid states such as NY and AZ.
CRITIQUE OF NODEXL

NodeXL is indeed a powerful toolkit that consists of a wide variety of handy tools, including importing tools, pre-processing tools, clustering tools. Besides, the whole process of exploring the networks could be automated. User just needs to populate some options. With a good idea of what to do and a sufficient amount of familiarity with the tool set, it is easy and fast to come up with a good visualization of networks, yet still has enough flexibility on almost everything in a graph such as coloring, positioning, widths, arrangement, etc.

Although it’s due to Twitter instead of NodeXL, and it’s unfair to blame it on NodeXL, I still had a bad experience when using NodeXL to import Twitter data due to the Twitter rate limits. Part of the reason for bringing it up is that it seems that the import tool is unreliable when it tries to resume from a mandatory pause. Another problem that made it less pleasant to use NodeXL was that when I’m trying to use the automatic column fill function, it is hard to get to a specific advanced option window, because as I click on any of the auto-fill entries, it selects/unselects that option, although what I really wanted was to see its advanced menu. I would suggest that the buttons for advanced menus to be made as separate buttons in the table auto-fill options.

Another issue is that the snapping of the graph window appears to be not so welcome. It seems that it’s something that’s hard to customize as a plugin of office, but if at all possible I would think an extra button for turning on/off the snapping of windows would be useful.

Since some times the documentation for certain feature is hard to find (e.g. the option to specify whether the group coloring overrides vertex coloring), I would suggest to have a wiki with better documentation to help the learning process.
REFERENCES

# APPENDIX

<table>
<thead>
<tr>
<th>Media Outlet</th>
<th>Twitter User ID</th>
<th>Avg. P Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fox News</td>
<td>foxnews</td>
<td>0.26</td>
</tr>
<tr>
<td>Washington Times</td>
<td>washtimes</td>
<td>0.29</td>
</tr>
<tr>
<td>Wall Street Journal</td>
<td>WSJ</td>
<td>0.41</td>
</tr>
<tr>
<td>USA Today</td>
<td>usatoday</td>
<td>0.51</td>
</tr>
<tr>
<td>CBS News</td>
<td>CBSNews</td>
<td>0.53</td>
</tr>
<tr>
<td>Washington Post</td>
<td>washingtonpost</td>
<td>0.54</td>
</tr>
<tr>
<td>Today Show</td>
<td>todayshow</td>
<td>0.55</td>
</tr>
<tr>
<td>ABC World News</td>
<td>abcworldnews</td>
<td>0.57</td>
</tr>
<tr>
<td>CNN Breaking News</td>
<td>Cnnbrk</td>
<td>0.58</td>
</tr>
<tr>
<td>The New York Times</td>
<td>nytimes</td>
<td>0.60</td>
</tr>
</tbody>
</table>