Find BING users in Twitter

- Jianyu Li  lijianyu9@gmail.com

Introduction

Google seems to be the most powerful search engine that used by the most of people in the World everyday. One of its major competitors is Bing. I am curious on people interested in talking Bing in Twitter network, and want to get insights on those who would be most influential in passing information around. Therefore, I explored the seed users who has tweets including hashtag “#Bing”, then expand seed users’ retweeters who retweeted their tweets in history.

Dataset

I used a dataset that was collected between April 25, 2011 and June 25, 2011. I firstly obtained all seed users who had at least one tweet including “#Bing” hashtag, then explored users who have retweeted seed users’ tweets at least once during this period. I chose retweet network for mainly two reasons. For one, retweeting is an influential and powerful way of diffusing information. For another, the retweet dataset may not be too large as followers network. Finally, I have a dataset with 727 twitter users.

Headlines

**Headline 1: Who is (are) the CENTER(s) talking about Bing?**

After importing the network to NodeXL, and choosing the Layout Frunchterman-Reingo, I got a basic graph. Then I calculated graph matrix under the Analysis, Graph Metrics, to get every user’s Betweenness Centrality, PageRank, Clustering Coefficient scores and so on. After that, I used Autofill to assign Betweenness Centality to Vertex Color, Vertex Shape and Vertex Size, and also set the labels and label positions. Besides these, I also used graph options to set the colors of vertex, edge and background.
From the graph, I could easily figure out that there are three big-sized light green vertices, whose Betweenness Centrality scores are the highest. From the label, it is easy to identify their user screen names in Twitter. By looking at their Twitter page, I found out that they are all interested in social network, photo graphics, and they had a huge number of followers. Bing search has nice photos on every day's home page, and it integrated social features when searching. From these users’ interests, it is no hard to find out why they tweeted about Bing. Using NodeXL, we could easily find the CENTERs who are talking about Bing.

**Headline 2: Number of Retweeter & Mentioner positively correlated!**

I am interested in exploring patterns between number of Retweeters and Mentioners for the users in networks. Therefore, I added two columns Mentioner Count and Retweeter Count to the Vertices worksheet. To draw a X-Y plot, in Autofill, I specified Vertex X as Retweeter Count, and Vertex Y as Mentioner Count, then used a logarithmic mapping in both vertex X/Y option, and also updated the Vertex Shape to be circle, and keep the Vertex Size to be determined by Betweenness Centrality. Furthermore, I also changed the option of Vertex Label Position to “Greater than 20000” then display labels.
From the X-Y plot, I figured out three interested things. Firstly, People with more Retweeters tend to have more Mentioners. We could see this pattern by following the average dots heights in Y-axis (Mentioner Count). Secondly, the three big circle users are all having a high number of Retweeters and Mentioners, which indicates that high Betweenness Centrality users usually have a fairly high number of Retweeters and Mentioners. Lastly, there are some outliers. Around 10 users with modest Betweenness Centrality score (<20000, because of no label display) got the biggest number of Retweeter and Mentioner Count. One user, the rightmost one, who got the biggest number of Retweeter Count, but a modest number of Mentioner Count.

**Headline 3: “CENTER users” has more power within his/her group, and across wider groups!**

I used group features tried to explore influences across groups. I ran the cluster algorithm under the Groups, Group by Cluster. Then I changed the Layout option to layout each of the graph’s groups in its own box and sort the boxes by group size. After that, I changed Layout to be Spiral.
From the graph, we could easily get two insights. Firstly, high Betweenness Centrality users got wider impact across groups. This could be seen from the edge density between groups, that the left most group center got the most retweets from other groups, and the retweets amount decreased from left to right. Secondly, the bigger the size of Spiral representation in each group, the higher the number of people that would be impacted by the center user.

**Conclusion**

Suppose Microsoft wants to attract more users to use Bing search by utilizing Twitter, picking the users with highest Betweenness Centrality would become a natural choice, because he/she will impact more people in both his/her own group and across other groups, and he/she would have a fairly high number of Retweeters and Mentioners. However, known from headline 2, picking some outliers with the high number of Retweeters or Mentioners would also be fairly good, because retweet and mention would have more impact in diffusing information.

**Tool Critique**

**Good things:**

NodeXL is powerful on its network analysis and the image visualization. It provides the group matrices such as PageRank, Betweenness Centrality, and clustering methods, which are basic matrices used to analyze social media. Besides, NodeXL gives users multiple choices on visualization options. Users could pick different layout, merge and expand groups, dynamically filter out non-interesting vertices, have box display for different groups and so on. Furthermore, NodeXL is flexible in functionalities. It allows users to add customized columns to different sheets, and allows different kind of data to be imported. Lastly, it is easier for experienced Excel users to get started because NodeXL is built upon Excel.

**Bad things:**

I think most of bad things are related to visualization options. The first thing I really do not like is the missing of UNDO. There is no UNDO button in NodeXL. When
AutoFilling a column, if I want to change it later, I have to go back to AutoFill, or manually delete the column, because there is no UNDO available. Also, for the graph window, suppose I want to change back to a state of my last graph, there is no way but only to change the settings to the previous state again. Besides, there is no way to import multiple worksheets together to the new worksheet. Further, the graph is so sensitive to be trapped into the worksheet, and it would feel better to drag graph window around freely.

The other thing I encountered is, there is no way to change the label color when tried to make label a different color from the color specified in Vertex Color in AutoFill using Graph Option. For example, in headline 1, the label’s color is accompanied with the Vertex color as indicated in Vertex Color in AutoFill. Suppose now I want to change the label color to be black in order to make it more visible, I tried to set it in Graph Option, then Label, then Font, but it did allow me to change it.