Twitter Sources and Sinks

Nathan Matthews
nmatthe2@illinois.edu

1 Purpose

Some people who use the Twitter microblogging service are quite prolific, posting dozens of status updates per day. Often these posters go beyond simple status updates to report local news, coordinate to accomplish tasks, carry on conversations, form groups, and take polls. Other users of Twitter hardly ever update their own status, instead using the service as a way of keeping track of their friends or other interesting people. Twitter can be used as a sort of alternative news feed or link blog. Some web applications use twitter as an easy alternative to the traditional API, and don’t perform any status updates at all. The purpose of this visualization is to graph social networks formed by friends/followers and to determine who on the network are sources of information, and who are sinks.

2 Visualization

The social network is ego-centric beginning with a single user and their friends and followers. Depending on the rate limit imposed by the Twitter API, the data could extend to two or more levels of links. The strength of a link is determined by several factors:

- how long the link has existed
- whether the link is mutual, i.e. both users following each other
- whether either user has marked the user as a favorite
- the number of cross-references made in tweets

In the visualization, users are represented by colored circles surrounded by bands. The size of the circle represents activity as measured by average tweet frequency. In the case of “super-source,” people with more than a set number of followers, the circle is enclosed by a band whose thickness indicates the number of followers. The same procedure applies for users who are “super-sinks” and follow many tweets. These are likely to be automated services. If a super-node repeatedly references another user in tweets, that user is explicitly shown, as for the friends and followers of non-super-nodes.
The hue of a node is determined by the relative number of friends and followers. Users who have more friends than followers, or "sinks," are shown as more blue. Users who have more followers than friends, or "sources," are shown as more green. These colors could be configurable. The saturation of the node's color is controlled by the extremity of the source or sink; a higher disparity is shown as more saturated. Partly this is to draw attention to their nodes, and partly because automated services often automatically follow every person using them, which would make their nodes grey.

A simple interactive element would be for mousing over or clicking nodes to display the name of the user and perhaps their latest status update. Clicking a link could display the strength of the link and a short statement of the reasons (age, favorite, number of references).

3 Implementation

Hopefully, the Twitter team will lift the rate limit so that a large amount of data can be gathered. I hope to get information out to at least two levels of connection. For this purpose, the Twitter REST API will be used. To search tweets for references, the Twitter Search API will be used, which has no rate limit. The twitter gem for the ruby programming language will be used for simplicity. User names, tweets, friends and followers will be stored in a database. The following table shows the colors of nodes based on their relative number of followers and friends:

| followers ≪ friends | blue  |
| followers ≈ friends | grey  |
| followers ≫ friends | green |

More precisely, if \( \alpha \) is the number of friends and \( \beta \) is the number of followers, and \( \gamma \) is as configurable brightness value, then the hue (H), saturation (S), and value (V) of the node are determined according to:

\[
H = \frac{\alpha \cdot \text{blue} + \beta \cdot \text{green}}{\alpha + \beta}
\]

\[
S = \frac{|\alpha - \beta|}{\alpha + \beta}
\]

\[
V = \gamma
\]

The strength of the links between nodes is used in their placement by creating a system of springs whose natural length is shorter for stronger links, making strongly linked nodes appear closer together. A standard layout algorithm could be used to get the minimum link crossings. The links themselves are drawn as arrows with one or two heads depending on who follows who. The size of the circles are determined by the equation \( d = 10 \cdot (\log_{e} f + 7) \) where \( d \) is diameter.
(in pixels) and \( f \) is in update frequency (posts per day). The rings have a thickness equal to \( w = 10 \cdot \log_{10} n \) where \( w \) is the ring width (in pixels) and \( n \) is the number of followers or friends.

4 Example

The example shown below is only partly based on real data; the center node is me and is colored correctly. The rest of the graph is not based on real data, but shows how the color, connections and rings would work. The graph below shows less nodes and connections than would probably actually be present at this depth of connectedness.