Who cares about the HCIL?

Panagis (Pano) Papadatos
squawk@umd.edu
November 8, 2012
CMSC734

Data Collection

The network of all the users that follow the University of Maryland Human Computer Interaction Lab’s Twitter account (Twitter handle “hcil_umd”) and their relationships (1.5 Egocentric network) were retrieved by using the “Import Twitter User’s Network function” that is provided by NodeXL. Follower/following relationships were examined, without investigating actual tweet content. In this network, nodes represent Twitter accounts that are following the account hcil_umd on Twitter. Edges between the nodes represent a ‘follows’ relationship between two people. The graphs produced are directed due to the directed nature of the ‘follows’ relationship. The node ‘hcil_umd’ and all its edges were manually removed from the network, since their existence is redundant due to the nature of the network; this is common practice for the study of egocentric networks. Clusters were automatically detected using the Clauset-Newman-Moore clustering algorithm.

The NodeXL File (including the raw data) is available here.

The Network

To this date (11/5/12) The network of people that follow the University of Maryland Human Computer Interaction Lab’s Twitter account has 196 nodes and 1690 edges. On a first look, this is a very sparse graph, since its density is 0.044 (only 4.4% of the possible edges exist). However, the data suggest that this network possesses small world qualities, since it has a relatively large average clustering coefficient (0.5) and a small average geodesic distance (2.2). This suggests that between any two people A and B that follow the HCIL on Twitter one of the following is likely to be true:

- A follows B or B follows A
- A follows/is followed by C, B follows/is followed by C

Furthermore, the reciprocated Edge ratio is 0.53, which means that 53% of the follows/is followed by relationships are reciprocated.
1. HCIL Pillars: Ben Shneiderman, Ben Bederson, Allison Druin

This graph shows all the Twitter accounts that had at least 4 followers within the HCIL network (in-degree ≥ 4). The color (and positioning) of each node signifies the cluster that node belongs to (detected by the Clauset-Newman-Moore algorithm). The size of each node represents its betweenness centrality. The layout algorithm used was Fruchterman-Reingold.

This graph shows that the 3 most prominent (and in-between) nodes in this Twitter network are Ben Shneiderman, Ben Bederson and Allison Druin. These results seem to be consistent with the results in Gove and Chanhyun’s project; in the HCIL co-authorship network. However, one very important ‘pillar’ seems to be missing. That pillar is Catherine Plaisant. This result was peculiar and interesting. Shortly after visiting Cathrine Plaisant’s Twitter page, we realized that her account is currently not following the HCIL Twitter account. This, of course, is not to suggest that she is not important within the HCIL; the fact that she only has 14 Tweets suggests that she is not an active user of Twitter.

Most of the clusters that were automatically discovered seem to have significance in terms of non-network related qualities. All of the nodes in the dark blue cluster is either an acclaimed HCI Professional or HCI Faculty (data retrieved from LinkedIn). Similarly, most of the dark green nodes are HCI students and some are HCI Faculty. The light blue cluster is mostly organizations (University of Maryland related and few others) and some Faculty and the light Green cluster does not seem to share a common theme. Lastly, it is interesting to observe that this looks like a very well connected network.
This graph shows all the Twitter accounts that follow the HCIL on Twitter, that have at least 500 followers. As in the previous example, the nodes are colored by cluster. The size of the nodes represents the number of followers that they have on Twitter and the layout was initially decided by the Fruchterman-Reingold algorithm but the nodes were manually adjusted to avoid label overlap. It is interesting to observe that the most influential people that are interested in the HCIL are not the ones that are most relevant to it (the “Pillars”). 3 people stand out in this graph (If we disregard the UMD Twitter account): Joyce Valenza, a renowned librarian with a stunning 13581 followers on Twitter; Miguel Rios, a Data Visualization technology leader that works for Twitter, with 5344 followers; Marc Smith, a chief social scientist that has heavily contributed in the development of the open source software NodeXL, which was used for this project. Ben Shneiderman, Ben Bederson and Allison Druin are still part of this graph, but they are not prevalent.
3. Most of the Low Tweeting Population within the HCIL network is not well connected and is not following a lot of other accounts

This graph shows all the Twitter accounts that have at most 150 tweets total; the nodes are sized by the number of people that they are following and colored by cluster, like earlier. The graph was initially laid out using Fruchterman-Reingold but the nodes' position was changed manually for clarity. It is interesting to observe that most of the nodes in this network are not connected to each other and follow a low number of people. However, one group that is highly connected in this category is a group of HCI students at the University of Maryland. One node that really stands out in this network is Brian Butler, who is faculty at the university.

NodeXL Critique

NodeXL is an extremely powerful tool that offers an impressive amount of customizability. A feature that I found really useful was the option to dynamically add columns and change any aspect of the data based on them. Other tools, however do not allow for this sort of dynamic environment since data usually has to be edited in different software and then reimported. However, the images produced by NodeXL are not by default aesthetically pleasant compared to other tools. This disadvantage is made worse by the fact that labels are very often
overlapping even when there is enough space for that not to happen. It would be really interesting to see a new version of NodeXL where the default images produced look smoother and the labels don't overlap when there is space. Another disadvantage is that NodeXL is very platform specific, and even though it is open source, it only operates with Microsoft Excel, which is proprietary and restrictively expensive for certain users. Lastly, NodeXL has a lot of inexplicable bugs, some of which are consistent (like group boxes appearing empty) and some of which are not (Filters disappearing or the program crashing).

Related works