Exploring Facebook Friendship Network with NodeXL

Peixin Gao
gaopei@umd.edu

Introduction

The friendship network of a person on Facebook is related to his/her friendship in daily life, and can be treated as a union of daily friendship networks in different time frames. For example, you may not stay close with your high school classmates after you graduate from university, but you may still keep in touch with them via Facebook. And it is one of the reasons why Facebook is so popular, to help get connected with people that may not be around you.

One interesting question about the friendship network on Facebook is how do your friends on Facebook interact with each other? For instance, are friends of similar backgrounds (for example same school) friends of each other? Do they form a tight cluster or a more relaxed or random structure? Do friends from different groups have connections? Are the connections strong?

To find out the answer to the question, I decided to explore my friendship network on Facebook using NodeXL, is a tool for social network analysis that can import data from online social networks like Facebook and Twitter.

Data

I installed NodeXL and its Social Network Importer, and used the importer to import data about my personal network via connecting my Facebook account.

The imported data form an undirected graph of my friendship network on Facebook, with 404 vertices and 9486 edges. In the graph there are 4 connected components among which three are single vertices. The diameter of the graph is 6 and average geodesic distance is 2.45.
I grouped the vertices in the graph into clusters using the Clauset-Newman-Moore cluster algorithm and laid out the graph using the Harel-Koren Fast Multiscale layout algorithm. The colors and widths of edges are based on edge weight values (which are all set to be 1). The vertex sizes are also uniform in the graph.

Then based on this graph, I used dynamic filter to filter out vertices of low degree (<54) and make them of 10% opacity, to generate a graph of high degree vertices.

Finally I collapsed all the groups and applied the average geodesic distance as the measure to have a collapsed version of the graph.

From the graphs generated by NodeXL, I found three headlines.

(1) From the first clustering graph, I found 6 groups, 5 of them are connected and one has only unconnected single nodes, and the 5 connected clusters in the graph are amazingly in accordance with the 5 main friendship circles that I have along my time lines, with high accuracy.

The biggest group has almost all the friends I met when I was an undergraduate. The second one is my current friend circle as a graduate student at University of Maryland. The third one consists of my friend in Japan. In the fourth one there are my classmates of high school. The fifth one has my friends in industrial design class. The last one has all singletons including me. The single vertices are the guys that I don’t really know. And since each edge means the friendship between two of my friends, so I am the one of single edges in the last group.

Thus I labeled the 6 groups as Undergraduate, Graduate, Japan, High School, Industrial Design Group and Outliers respectively. The graph corresponding to this headline is as follows.
(2) From the filtered graph, I found that the first group, i.e. the group of friends during undergraduate period, is more closely connected than the other groups, means that friends of mine are most likely to be friends in this group.
(3) From the graph of collapsed groups with the average geodesic distance, I found that the first group is the “hub” in the whole network. Every group has a short distance to it than other groups.

Critique

Based on my using experience, NodeXL is very powerful in many aspects. First of all, it is a template in MS Excel and easy to adapt to and use with no programming skills required. Many operations are simplified and can be done within several clicks. It is also very easy and flexible to import data in NodeXL with direct connections to social networks. There are a lot of options for different types of network visualization and analysis. Many parameters are configurable and many calculations (e.g. metrics calculation and grouping) can be done within it. I like the dynamic filtering because it helps me focus on vertices of important properties like degree and find insights from the graph.

Besides, I feel that it would be better if the tool is more reliable, since it froze many times during my work. Also, I guess some more functions are needed. For example, currently I found that NodeXL doesn’t support flexible group labeling (maybe I was wrong), so in order to label the groups, I had to choose “label: Top Domains in Tweet” and input what I want to label in the “Top Domains in Tweet” block. Also, in the collapsed graph, I wanted to make the size of the vertex related to the size of the group it stands for, but I didn't find a way to do it.