CSC443 Assignment 1.1 and 1.2

Renata Boodram (boodram8, 999872320)
Tyler Pham (phamtyle, 999643382)

Part 3: Research: Twitter Graph Degree Distributions

Experiment 1: Power-Law Distribution

The purpose of this experiment was to determine whether the Twitter Graph displayed signs of conformism. In social networks, the number of of out-going links and incoming links often can be related to a power-law distribution. The experiment and its methodology will be described here.

Methodology

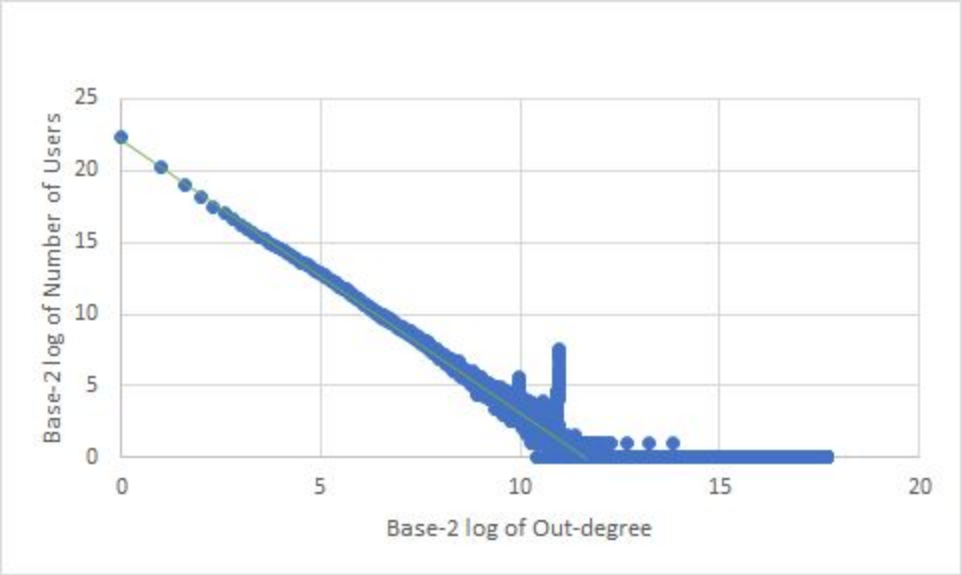
In order to determine the number of outgoing or incoming links for a particular user, the maximum outgoing or incoming degree was first calculated. Maximum degree refers to the maximum number of outgoing or incoming links that are observed in the provided Twitter Graph dataset.

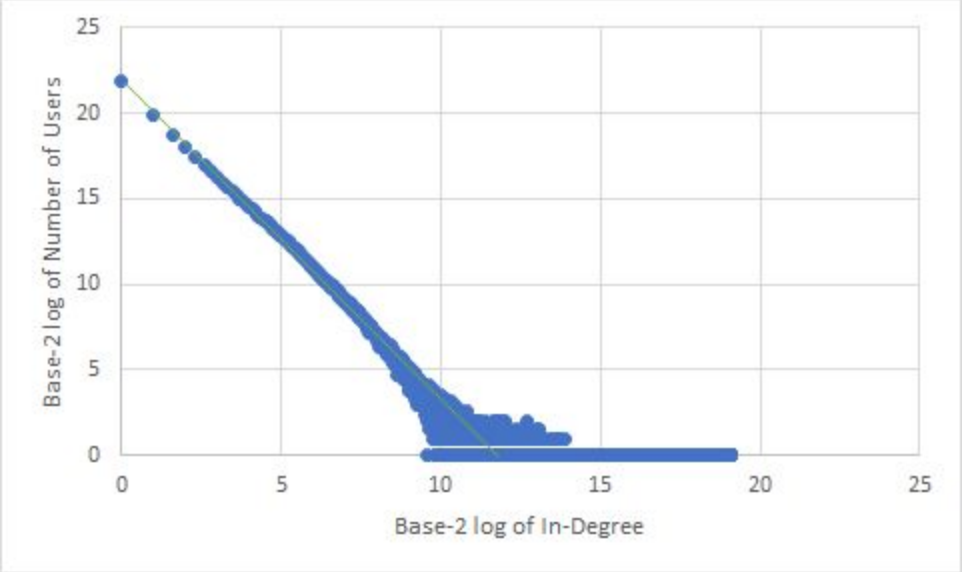
When calculating outgoing links, we used a file sorted by the first user ID where each line is tuple of two user IDs stored in the file. The relationship of the tuple involved the first user ID having an outgoing link to the second user ID; naturally, this means that the second user ID has an incoming link. So, when calculating incoming links, we used a file sorted by the second user ID.

To calculate the number of outgoing or incoming links, we read the sorted file until we encounter a new user ID, while keeping a count of the current user's link count. We only update the corresponding degree count when we encounter a new user.

Calculating the C value

As mentioned previously, the purpose of this assignment is to find the "c-value" in the power-law distribution formula, $f(k) = \frac{a}{k^c}$ where k is the number of links and f(k) is the fraction of users that have k number of in-degree or out-degree links (whichever applies to the current situation). Taking log of both sides, we can use the equation $\log f(k) = \log a - c \log k$ to isolate for c. This is what we did for both in-degree and out-degree which produced the graphs below:





For out-degree, we determined that the trendline that best fit our results is the equation $y = -1.838065276x + 22.38039474$ where the determined value is $c = 1.83806527$.
Similarly, for in-degree we determined the trendline that best fit our results is the equation: $y = -2.19076x + 21.84873487$ where the determined value is $c = 2.19076$.

Conclusion

Based on our values chosen above, it appears that both the out- and in-degree follow a power-law distribution, indicating that they are a "scale-free" network. In-degree has a steeper slope than out-degree which indicates that the rate at which the number of followers for a user decreases more drastically than the number of users which a user follows.