# CSC443 Assignment 1 Part 2 Research Results

RESEARCH: TWITTER GRAPH DEGREE DISTRIBUTIONS

Experiment

We wrote a C program named "distribution" to get the counts (number of degrees) for in-degree (number of followers per UID) and out-degree (number of people a UID followers) for all degrees (from 1 to the maximum degree). The program outputs two text files (one for in-degree and one for out-degree) that each file contains two columns. The first column is the degree and the second column is the counts for the corresponding degree. The rows are sorted in an increasing order.

After running the C program, we import the two files generated by it to Excel, calculated the log of the degree and counts columns, and plotted the data using a scattered form diagram. The result we got is shown below:

Distribution of Out-Degree
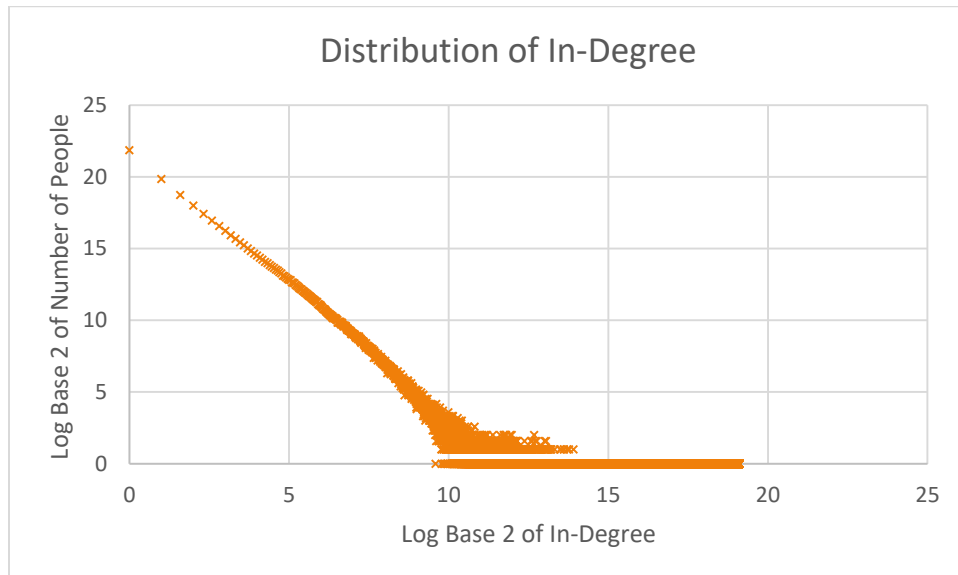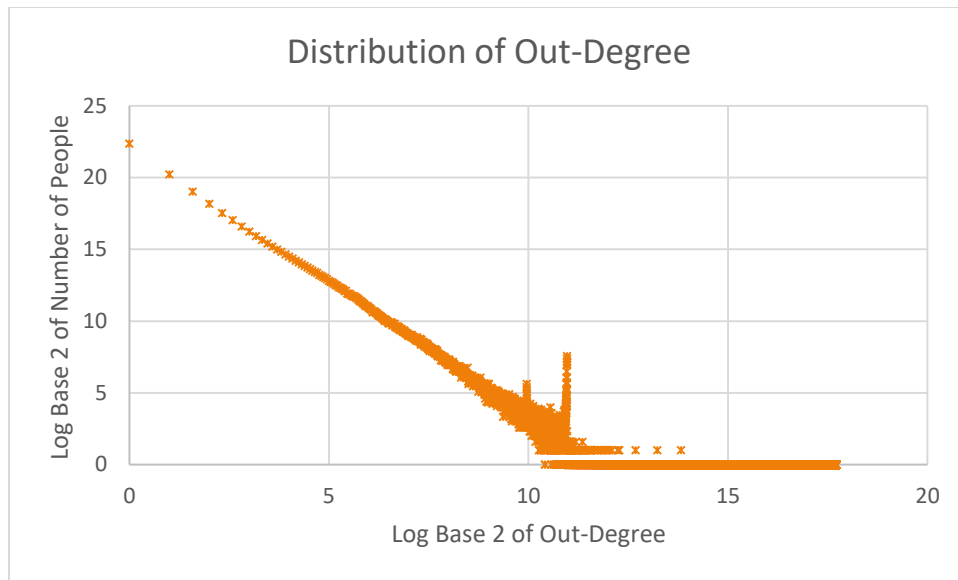
Conclusion

The diagram demonstrates that both in-degree and out-degree exhibit a power-law distribution. The larger the degree, the fewer the count. Also, the difference between the median and maximum degree is very large.

CSC443 Assignment 1 Part 3 Research Results

1 QUERY RESULT

Total number of true friends: 21776094

Total number of distinct users: 11316811

Top 10 Celebrities			
User Id	In-degree	Out-Degree	Difference
5994113	564512	292	564220
7496	350885	6035	344850
1349110	341963	1472	340491
1629776	172231	2120	170111
8121005	155967	34	155933
2041453	152689	620	152069
797152	118826	74	118752
6623784	116002	183	115819
645019	107914	275	107639
3403	102877	2946	97931

2 SUMMARY

True = Total number of true friends

Total = Total number of distinct users

$True/(Total^2 - Total) = 0.000017\%$

Therefore, there're just 0.000017% pairs of users on Twitters over all pairs of users are true friends. About 5% of all users are following the most popular celebrity who has UID 5994113. The question that I would like to ask is what the distribution of the difference of in-degree and out-degree would be like.