Brandon Patton

Online Social Networks

Assignment 9 – Analyze Social Network Data

I. Purpose

The purpose of this program is to analyze the given Facebook data and draw conclusions based on perceived correlations between certain attributes of data. This allows us to experiment with ideas as to why certain individuals have more friends than others, what influences friend count, what influences likes received, among others. By using the matplotlib plot functions, I was able to visually represent interesting comparisons of data sets and attribute these visuals to possible correlations between two variables.

II. Program Description

Much of this program was made with the help of matplotlib functions. Medians were received by the use of the statistics module. I opted to use medians rather than means since medians are unaffected by skewed data while means typically are; skewing of which would have significant impact on the representation of data.

I started by initializing empty variables for all possible data fields in the given csv file. I did not end up using all of them in my plots. I went through and converted all the string integers to integers in order to be used in plotting functions later.

Plot 1:

Plot 1 compares the total amount of friendships to the total initiated friendships using a scatter plot. The x coordinate for a given point is a user's total initiated friendships and the y coordinate for that same point is that same user's total friendship count. This comparison was interesting because it showed a significant trend deviation as total friendship count increased. After reaching about 2200 friends, much less users tended to initiate friendships as often as from 0-2200 friends. I think this may be because once users reached a large number of friends, they tended to stop adding so many new friends as they had already encompassed all the people they possibly could have known and more.

Plot 2:

Plot 2 compares the total amount of friendships to the total number of days on Facebook using a scatter plot. The x coordinate for a given point is a user's total number of days on Facebook and the y coordinate for that same point is that same user's total friendship count. This plot was interesting because it seemed like the longer a user had an account on Facebook, the less friends they ended up having. This could be due to an initial rush to increase friend count when creating a new Facebook account and realizing over time having such a large amount of friends is hard to manage/unrealistic to the amount of people they actually communicate with.

Plot 3:

Plot 3 compares the total amount of friendships to the total amount of likes received using a scatter plot. The x coordinate for a given point is a user's total amount of friends and the y coordinate for that same point is that same user's amount of likes received. This was an interesting comparison because it showed that the relationship between friends and likes received did not matter at a certain level of likes. The uselessness of this graph could be mitigated by not accounting for the outliers that caused the graph to measure to such high numbers. However, it seemed at certain levels of friends (1000-2000 and 3500-4750) likes seemed to increase substantially compared with other areas.

Plot 4:

Plot 4 is a bar plot that displays the median likes received by a given gender. The first bar represents the median likes received by Males, the second bar being Females, and the remaining bar being likes received by those identifying as N/A. This specific plot was very interesting to me since it displayed such a high disparity between the likes received by males and females, with females receiving just under 30 median likes while men receiving just under 5 median likes. It is also notable that those who filed gender “N/A” received almost an exact middle amount between males and females.

Plot 5:

Plot 5 is a bar plot that displays the median number of friends by ages. User ages are displayed at the x-axis of the graph while the median number of friends for that given age are displayed on the y-axis. This plot was interesting to see how the number of friends different for users for each age. It seemed that young users (before and a little after age 20) had in general more friends than those who were between ages ~20-85, though those past age 85 displayed a massive increase in their number of friends, especially those who made it close to or above age 100. This is interesting because this plot displays a population of users who have significantly less friends than those at the beginning or end of their life. It would be interesting to find out why those closer to the age of 100 increased their Facebook friend count so significantly, even when compared to younger users that are expected to use the platform more.

Plot 6:

Plot 6 is a bar plot that displays the median number of likes by ages. User ages are displayed at the x-axis of the graph while the median number of likes for that given age are displayed on the y-axis. Although likes and friend count are probably unrelated for the most part, it seems like this plot is somewhat the opposite of what I would have expected from the previous plot. It seems to show that the number of likes peaks at around 3 different ages: just before 20 years old, around mid-50’s, and around age 100. I would have assumed otherwise that younger users would be receiving more likes simply because I would also have assumed that they are more active on the site in general. This apparently is not the case as those in mid-life and those who are around age 100 receive much more likes than younger users, the latter of which receiving astronomically more than any other age group. Once again, those who are closer to the end of their life seem to be much more active than I would have expected give the typical thoughts around social media literacy as it relates to age.

III. Observations

It was interesting to see the relationships between likes and gender as well as likes and age. My initial thoughts on likes and gender were that in general females would receive more likes than males, though I had no real predictions on those identifying as N/A. This hypothesis was realized for the most part as females ended up having a significant number of more likes over males. Additionally, those identifying N/A showed more likes than males as well, but almost exactly half of those received by females. My initial prediction for age as it relates to likes and friend count was based on the idea that younger individuals are more active on Facebook in general. This latter idea on activity seemed to be shown to be partially false on the friend count vs. age plot, as younger users beat those in mid-life but in the end were beaten out by those around age 100. As for median likes received vs. age, my initial thoughts were again that younger users would show far more than other ages. This hypothesis was also partially false, but out of the 3 age groups that demonstrated peaking like amounts, younger users seemed to be the least. Those around mid-50’s demonstrated higher median likes than those younger and in a wider group. Both groups were astronomically beat out by very old users around the age of 100. So, although my hypothesis that younger users would produce higher median likes was not totally wrong in relation to all data, it was by far the least out of the 3 outlying age groups demonstrating high median likes. This is interesting because I think the perception of social media is that young people are using the platforms more often than those who are older, though this data seems to disprove that to an extent.