Homework Assignment -5

Nithish Raghunandanan
Matriculation No: 03667351

Analysis of the Distance & Friendship Probabilities

I chose the tie strength threshold for friendship indicator as 50.

I sampled the friendship probabilities in buckets of 50 kilometres as I felt that its a sufficiently small interval to allow for friendships based on geography and at the same time not have too many buckets.

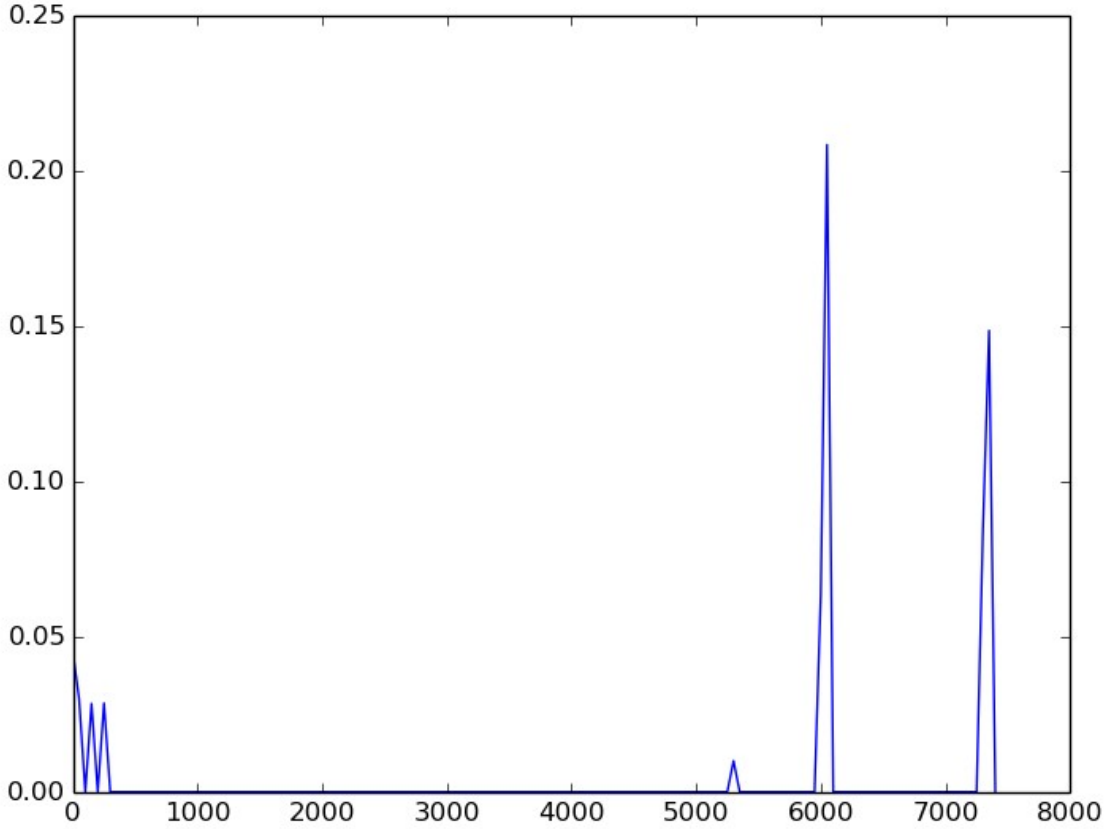

Illustration 1: Variation of the friendship probabilities(Y-Axis) with Distance in kilometers (X-Axis)

On analysing the probabilites, it can be seen that there are very few buckets with probabilites of friendship greater than 0. For example, there are friendships only in the buckets between 0-300km, then from 5300-5350, 6000-6100km and finally between 7300-7400km.

Comparison with Liben-Nowell et al.[1]

According to the paper by Liben-Nowell et al., the probability of friendship decreases with increase in the distance, i.e., the probability of friendship increases with geographic proximity. However, in our social network, it is seen that the statement holds for the initial few non zero probabilty buckets(upto 5350km). Then, it is seen that the probabilities are more than the initial buckets, i.e, the friendship probability increased with increase in distance. The result is similar in the variations of the probabilities even if the bucket widths are decreased to say 10km.

I would say that this may be due to the fact that we do not have enough samples in the

higher buckets to justify the statistical meaning of probabilities. The number of ties at the higher distance factors are lower compared to those at lower distances(See table below). It could be analogous to getting a Head on two tosses of a coin and saying the probability of getting a Head as 1. The number of frienships are more at the lower distances. The friendships at higher distances may be due to some other factor like doing the same course or having worked together sometime before. If the sample sizes are comparable, I would say that it would most likely follow the results of Liben-Nowell et al. This is similar to the factoring of population density as described in [1].

Table showing the number of ties in each interval.

Interval	Probabliity	No. of Ties
0.0-50.0	0.0450833617	5878
50.0-100.0	0.03	300
150.0-200.0	0.028436019	211
250.0-300.0	0.0285714286	105
5300.0-5350.0	0.01	100
6000.0-6050.0	0.0634920635	63
6050.0-6100.0	0.2083333333	48
7300.0-7350.0	0.0833333333	12
7350.0-7400.0	0.1485148515	101

References:

[1] Geographic routing in social networks by David Liben-Nowell, Jasmine Novak, Ravi Kumar, Prabhakar Raghavan, and Andrew Tomkins