Friendship - The Two-sided Tale

Altaf Shaikh, Nimrah Jawed, Syed Hammad Ali, and Arham Ahmed

Dhanani School of Science and Engineering, Habib University, Block 18, Gulistan-e-Jauhar, 25290, Karachi, Pakistan
as05016@st.habib.edu.pk, nj04707@st.habib.edu.pk, sa04324@st.habib.edu.pk, aa05179@st.habib.edu.pk

Abstract. Friendships, even though a mutual relation, has two sides to its tale. The report explores these two sides through friendship paradox and identifies how the people at the two opposite ends of the spectrum perceive the dynamics of the friendship and use it to evaluate themselves in the wider context. By employing a graph-theoretic approach and analyzing the structure of scale-free networks along with understanding phenomenon such as class size paradox and Lake Wobegon effect, the report breaks down the underlying reasons and implications of the friendship paradox. It also looks into the idea of economies of attention and how the number of friends correlate with the said idea. The report applies the referred ideas in the local context of Habib University and its students and studies their perspectives under the light of the aforementioned phenomenon.

Keywords: Friendship Paradox · Class Size Paradox · Scale-free · Lake Wobegon.

1 Introduction

In the contemporary world where human networks are ever-expanding both physically and digitally - the consequences of such networks are widespread. The social implications of these networks greatly affect the human mind and perspective. Friendship networks are one such example. Friendships do not just provide happiness, support and security; they also provide a means to humans to evaluate themselves and others. People do not generally have a number in their mind to judge whether they themselves have an adequate number of friends or not. Therefore, the reasonable measure that they find is their friends. If an individual resorts to such a comparison, they would likely find themselves relatively inadequate - or to have fewer friends than their friends. This idea known as "Friendship Paradox" was first put forward by Scott Feld in 1991. (Feld, 1991)

In this report, we would be looking at the mathematical logic that results in the friendship paradox. For this, we would be using a graph-theoretic approach along with statistical methods such as the mean and variance. Furthermore, we would be looking at the inevitability of the friendship paradox in light of scale-free social networks.

The report will also look at the implications of the friendship paradox in the form of "Class size paradox" and also contrast it with "The Lake Wobegon" effect". Additionally, we would understand the trade-off of friendships and attention in the form of attention economics.

Finally, we would be applying the above ideas and models to study the real-world applicability in a local context - students of Habib University. This would give us an in-depth perspective on the paradoxes and perceptions revolving friendship.

2 Background and Scope

In order to analyze the friendship paradox through logical reasoning, we would be using some statistical methods along with fundamental ideas of graph theory. Consequently, we would also look at how the friendship paradox is inevitable by understanding scale-free networks. Then, to understand the social implications that come with the friendship paradox, we would be looking at some psychological effects. Furthermore, we would understand the idea of attention economics and how that plays out in the context of friendships in the modern world.

2.1 Networks and Statistical Mean and Variance

Network theory involves the study of the way elements in a network interact using graphs. Such networks are highly significant when studying human social networks. We would be making use of them, particularly friendship networks (where each node is an individual and the edge connecting two nodes represents friendship).

When analyzing data sets at large, the two most important measures are the mean and the variance. The mean is the "center of gravity" of the data, and is meant to carry a piece of information from every member of the sample. Likewise, variance is also an important measure that is used to study the amount of dispersion and uncertainty in a data set. The mean and the variance are given by the following formulae:

$$\mu = \frac{\sum_{i=1}^{n} x_i}{n} \tag{1}$$

$$\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n} \tag{2}$$

where x_i are the sample values, μ is the mean of the sample and n is the sample size.

2.2 Scale-free Networks

Often the idea of degree distribution is used to capture information about the structure of a network. The degree distribution is a plot of the degrees on the x-axis and the number of nodes on the y-axis.

A scale-free network is a network whose degree distribution follows a power law distribution. This means that most of the nodes have lower degrees and fewer nodes have a higher degrees. The idea of "Preferential Attachment" is at the heart of such networks i.e. the likelihood of a node getting a new edge depends directly on its current degree k. Scale-free networks are often used to model human networks as these embody similar features. We would be using the light of scale-free networks to portray inevitability of the friendship paradox.

2.3 Class Size Paradox

Often universities use the average class size to convey their statistics to potential parents and students who would want to get admission there. However, the average class size at a university conveys little about the experience of the average student there. Unless all classes have the same number of students, the average student will always experience larger classes than the school's average, simply because large groups have more people in them. This idea is known as the "class size paradox".

This idea is very relevant to many different social aspects where the experience of a person is different from the actual facts and figures. It is because of this phenomena that people often experience parks and beaches to be more crowded than they really are.

2.4 The Lake Wobegon Effect

The Lake Wobegon effect, a natural human tendency to overestimate one's capabilities, was named in honor of the fictional town created by Garrison Keillor as the setting of the "News from Lake Wobegon" segment of the radio program A Prairie Home Companion. The show ended with the monologue "Well, that's the news from Lake Wobegon, where all the women are strong, all the men are good-looking, and all the children are above average." This phenomena arises due to self-serving bias. Most people think that they are above-average drivers. The main idea of this phenomena is that humans tend to use the small sample in their experience as representative of the wider population. For example, a particular individual would only have experienced other people driving to a certain extent and using that little sample the individual begins to believe that they themselves are an above average driver in the context of the wider population.

Often people employ a similar bias and are affected by the Lake Wobegon effect when they use their friends as the sample to evaluate themselves.

2.5 Attention Economics

Attention economics is an approach to the management of information that treats human attention as a scarce commodity. The attention one can pay to friends decreases as their number increases. At the same time the amount of information that is available to a user increases as the number of friends increases.

3 Previous Applications (Literature Review)

3.1 Scott L. Feld and the Friendship Paradox

In his famous publication "Why your friends have more friends than you do?", Feld explored the Friendship Paradox for the first time and used networks and statistics to mathematically explore the phenomena(Feld, 1991). He used the book *The Adolescent Society* by Coleman (Coleman, 1961) to collect data about friendships among the students in 12 high schools. Feld initially restricted his analysis to 8 girls from Marketville High school and then moved on to the larger student body at the said school. Following is the friendship data of those 8 girls:

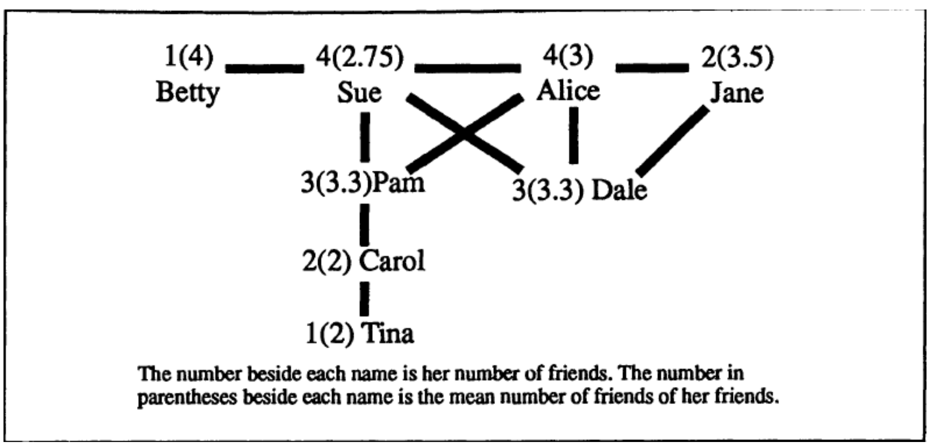

Fig. 1. Friendships among eight girls at Marketville High School

Feld compared the mean number of individuals' friends with the mean number of friends' friends. For most of the girls their number of friends was lower than the mean of their friends' friends. As can be seen in the figure, except Sue and Alice, all other girls have mostly less number of friends than the mean of their friends' friends (Carol has equal). This means that only Sue and Alice would feel relatively advantaged and all others would feel relatively deprived and inadequate. The same pattern was observed by Feld among the boys in Marketville and among the girls and boys of the other high schools reported in the Coleman study.

Feld pointed out that there are two distributions at play here. The first being the mean of an individual's friend which is just the sum of the x_i divided by the total number of individuals. The second is the mean number of friends' friends where the individuals are counted as many times as they have friends. Feld goes on to show that mathematically this would be equal to: mean number of friends of friends = $\frac{\sum x_i{}^2}{\sum x_i}$ = mean(x) + variance(x)/mean(x) (Appendix A). Meaning that the mean number of friends' friends would always be greater than the mean number of an individuals friends.

We would be using a similar mathematical analysis to identify the friendship paradox among the students of Habib University.

3.2 Scale-free Networks Applied to Twitter

In the 2015 paper published in Entropy (Aparicio, Villazón-Terrazas, & Álvarez, 2015) , scale-free networks were applied in the context of Twitter. As part of the study, the paper showed that Twitter can be considered a scale-free network with a small-world property. Data of about 50 million users was recorded and used for the purpose of this study. The specific model used was a generalization of the Albert-Barabasi network. Using the data of the sample users, a power law distribution of followers was depicted in the results(Appendix B). Scale-free networks are often applicable in social (human) networks be it digital or physical. We would also be analyzing that the friendship network at Habib University follows a power law distribution or not.

3.3 Hemenway and the Class Size Paradox

As mentioned earlier, class size paradox is applicable in a number of different scenarios. The earliest of these was studied by Hemenway (David Hemenway, 1982). He conducted his study where his sample was the Harvard School of Public Health. There were a total of 111 courses offered at the school. The average class size in the administration's view was 14.5 students. The expected class size for a typical student was 78(Appendix C). This discrepancy was due to the existence of some very large classes. To be more precise, only three courses had students above the figure of 78.

3.4 The Lake Wobegon Effect and its inverse

Studies in the field of social cognition and social networks show that people often have biased perspectives regarding themselves and their attributes. Some common studies that support the idea of the Lake Wobegon Effect are that a grate majority of people believe that they drive better than the average person

(Svenson 1981). Most people think they are fairer than the average person. (Messick et al. 1985) and they have better health prospects than the average person (Weinstein 1980).

Similar to the above studies, Nan Maxwell studied the effect in the context of student self-reported data (Maxwell,1994). From a pool of 2600 students at 10 U.S. colleges and universities, GPA was overstated by 45.9% of the students and likewise the SAT scores were overstated by 34.1% of the students. The study showed that an average student does overstate their own achievements. We will use the Lake Wobegon effect and its inverse to study the idea of popularity among individuals from Habib University.

3.5 Attention in Friendship

In the 2014 paper "Friendship Paradox and Attention Economics" by Subhash Kak (Kak,2014), he explores the idea of attention among friends (. One aspect, among others, that he explores is the amount of attention that a person can give to their friends. He depicts this using the following graph:

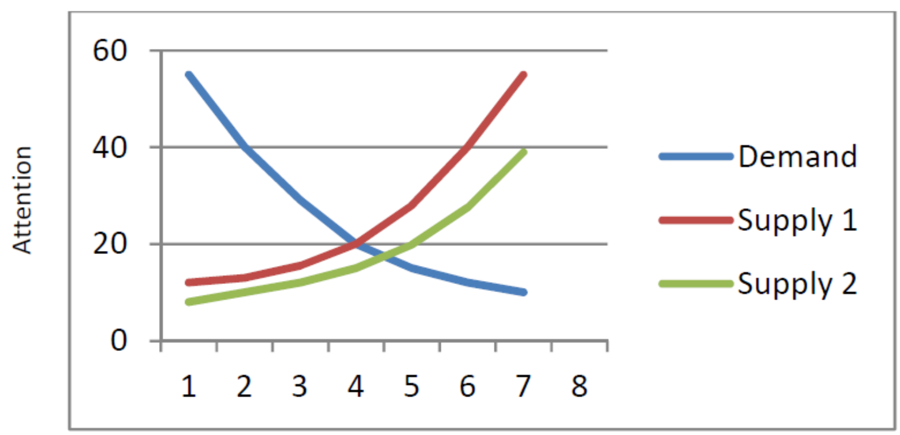

Fig. 2. How the attention (y-axis) varies with the count of friends (x-axis)

The blue line depicts the demand of attention towards friends that is met. The red and green lines depict the supply of information available to an individual. It can be seen that a person with less number of friends can give more attention to them compared to someone with larger number of friends. Likewise, the amount of information available to an individual increases with the number of friends.

Furthermore, there have been various studies on attention economics in terms of social networks. A recent study is performed in the paper "Attention Economy:

The Future of Social Networking Sites" by Gerard (2020). Although these mostly regard the idea in the context of how attention inevitably leads to revenue.

We, on the other hand, would be exploring the position of an individual in a friendship network with respect to the attention they get.

4 Application to a Contextualized Case Study

In order to explore the friendship paradox along with its causes and implications in a local context, we conducted a study at Habib University. We organized a survey that gave us valuable insights on the dynamics of friendship and the social psychology of the people involved.

4.1 The Setting of the Study

Our study was restricted to the class of 2022 of Habib University. We decided to restrict the sample of our study in terms of the batch rather than a particular major. The reason for this was that there would have been little variance in the type of data we would have acquired from friendships in a single major. Widening the scope to a full batch would incorporate all majors and would a much better interpretation. Out of a batch of 170+ students, 102 students took part in our study. These students were divided among the four majors at Habib University (CS, EE, CND, SDP).

4.2 Underlying Assumptions and the Network

There were some assumptions that we had to make when conducting this study. As friendship is generally a symmetric relationship, we got the inspiration from Feld's study and kept our focus on only those friendships which were mutual. Individuals were asked to name their friends belonging to the class of 2022. Only those pairs of individuals who chose each other were considered friends. That is how we constructed our network of friendships.

Another abstraction was regarding the quality of the friendships. Who does one consider a friend or not? Do they include only the close friends or even the people they are acquainted with? We asked the individuals to include all those people that they themselves would consider as friends. This, of course, abstracts away the quality of the friendships. This also helped us get an insight into one-sided friendships.

The names of the individuals were anonymized through numbers to keep the confidentiality intact. Following was the friendship network that was generated:

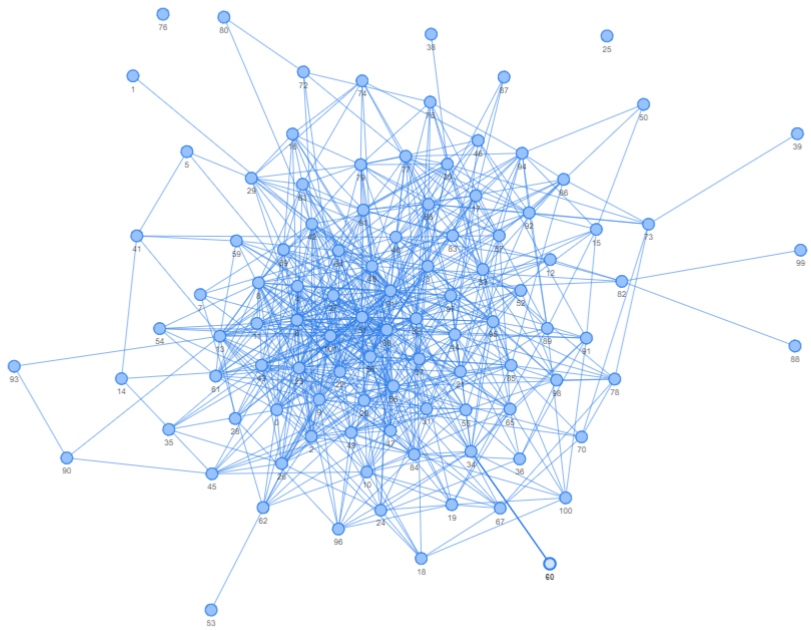

Fig. 3. Friendship Network of Class of 2022 - Habib University

The network remains disconnected because of two particular nodes (25 and 76) that have no connections. There are a total of 102 nodes (individuals) labelled from 0 to 101 and there are a total of 696 edges (friendships) among the nodes.

4.3 Network Properties

Following are the network properties of the network that we constructed:

Properties	
Property	Value
Global Clustering Co-efficient	0.408
Network Diameter	5
Network Density	0.141
Average Neighbours	13.92

4.4 Exploring Friendship Paradox

Initially we calculate the total number of friends for each individual in the sample and then divide it by the total number of individuals in the sample - this gives us the mean number of friends of an individual. Since there are 696 edges, total number of friendships are the double of it, 1392, since each edge contributes twice to the total number of friends. Therefore,

$$\text{Mean No. of friends of individuals} = \tfrac{1392}{102} = 13.65$$

Next, we would find the mean number of friends of friends in the network for each individual. Using an automated code by summing all friends of friends for each individual gave a total of 28498. Dividing this by the total number of friends, 1392, gives:

$$\text{Mean No. of friends of friends} = \tfrac{28498}{1392} = 20.47$$

As can be seen, the mean number of friends of friends is greater than the mean number of friends of individuals. Therefore, this calculation shows that the friends of an individual generally have more friends than the said individual.

Delving into this further, we also studied how many individuals would feel 'advantaged' or 'deprived' relative to their friends. This was done by comparing the the number of friends of an individual with the mean number of friends of that individual. We found out that from a total 0f 100 students (excluding the 2 disconnected nodes), 19 of them were relatively advantaged and the other 81 were comparably deprived.

4.5 Lens of Popularity

We have mathematically elucidated the existence of the friendship paradox. Now, we would try to see it through the lens of popularity which can be measured with the degree centrality of a node. We divided the nodes among three categories of popularity: Very popular (25 or more friends), mildly popular (more than 15 and less than 25 friends) and less popular (less than 15 friends). Following is the friendship network viewed from this lens:

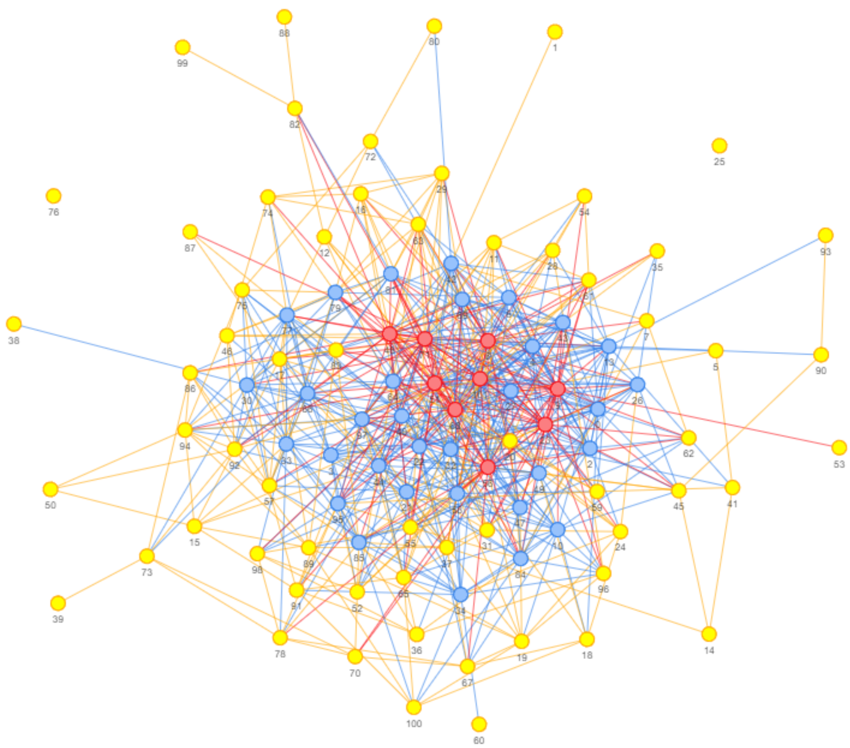

Fig. 4. Friendship Network of Class of 2022 - Habib University (viewed from a lens of popularity)

In the above figure, the red nodes are the most popular, the blue nodes are mildly popular and the yellow nodes are less popular. We can see that only 9 individuals fall into the category of most popular, 32 in the category of mildly popular and 61 in the category of less popular. This also gives us an insight into the centrality of the nodes in the network. The more central nodes in the context of degree centrality are towards the center. The less significant ones are towards the periphery of the network.

4.6 Degree Distribution

To see the overall distribution of how the degrees are structured, we would be plotting the degree distribution of the network. The lowest degree in the network was 0 (for the two disconnected nodes) and the highest degree in the network was 43 (for a single individual). The highest frequency of a particular degree was 7 - for multiple different degrees. The egree distribution of a standard scale-free network as shown in Fig.6 can be compared with the best fit curve of the overall network (Fig.5).

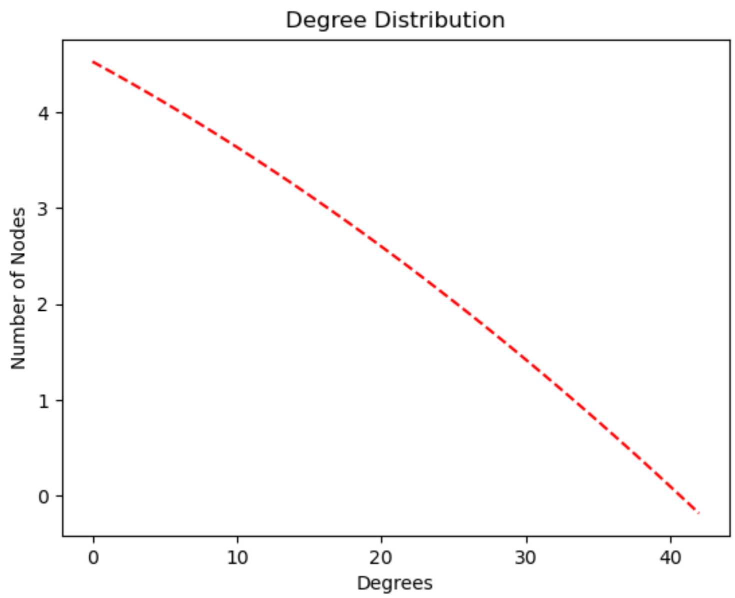

Fig. 5. Degree Distribution of the network - degrees (x-axis) and number of nodes (y-axis)

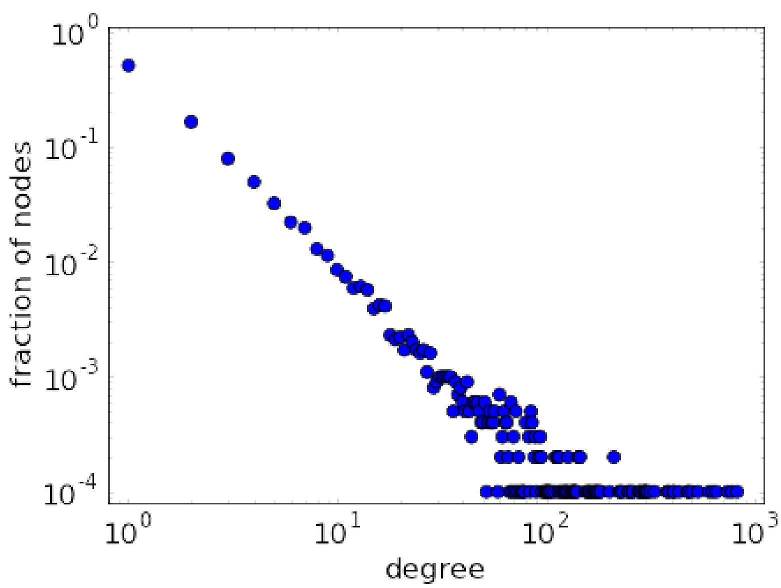

Fig. 6. Standard Power Law Distribution

We can see that the friendship network that we have for the class of 2022 at Habib University is very similar in its shape to the standard power law distribution. There are two things to consider here. In our case, we plotted the number of nodes on the y-axis while in the standard case the fraction of nodes with a particular degree are plotted on y-axis. This should not be a differentiating factor as both are just different scales of the numbers and hence show consistent depiction. The other aspect to consider is that the range of values of both the degrees and their frequency are much lesser in our case. This is in alignment to the fact that we had a smaller sample size (102) than the standard power law shown where degrees are up to 10^3. Again, this is a factor of the sample chosen and should not affect the overall results. It is, however, worthwhile to note that the decline in the frequency of the degrees is not as sharp in our case as is in the standard power law distribution. This is due to the fact that the variance in the frequency is much lesser in our case and hence the decline in the frequency is not as mathematically and visually sharp.The existence of scale-free networks does make sense in the context of university as individuals often prefer to be friends with people that are popular - preferential attachment. This creates hubs in the network.

The trend we got from the sample is along the lines of a scale-free network. However, as the data is taken from a certain category of people i.e. students belonging to the same batch at Habib University, the variance is lesser and obviously shows up in our results. This also goes on with the fact stated in the recent paper "True scale-free networks hidden by finite size effects" (Serafino, et al., 2021)that scale-free networks can be used to model human social networks but due to finite sample sizes, they may not always strictly follow the power law distribution.

4.7 The Social Implications

As part of our survey, we asked students two questions that focused on their perception of their own popularity and their ability to make friends. Following are the results of the two questions:

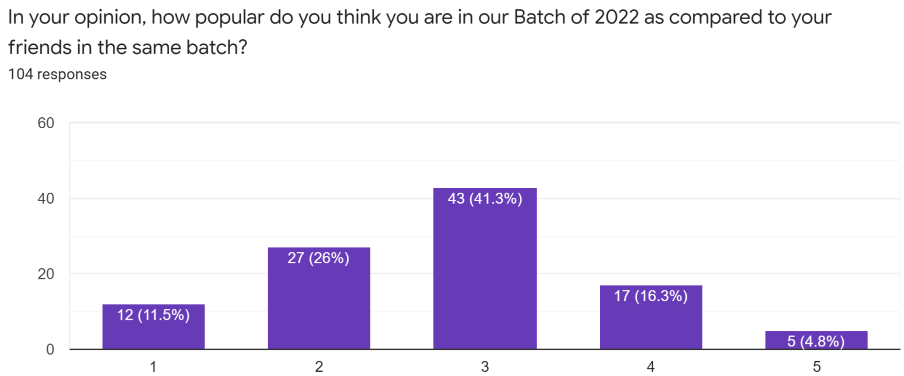

Fig. 7. Perception of Popularity (5 being Very High)

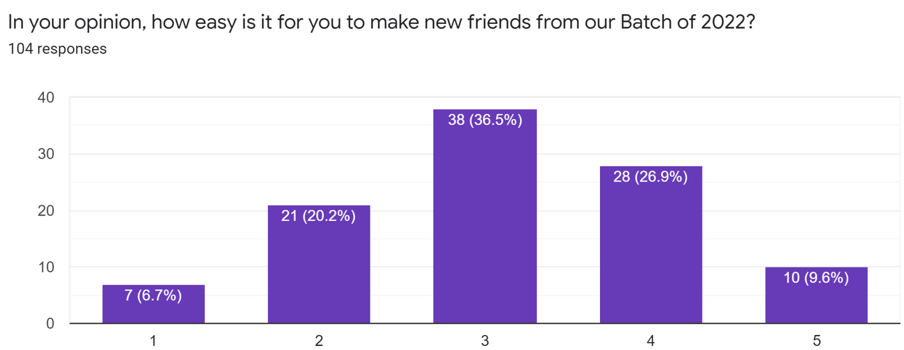

Fig. 8. Perception of ability to make friends (5 being Very Good)

Looking at figure 7, we can see that 37.5% of people chose 1 or 2 and 21.1% of people chose 4 or 5. The majority, 41.3% of people, chose 3. This implies that lesser people think they are very popular and most think they are mildly popular.

If we were to look at this from the perspective of class size paradox, most people should feel that they are less popular and a smaller number of people should believe that they are highly popular. In our friendship network, less people put themselves in the 'most popular' category, and more people put themselves in the 'less popular' category. However, the wide majority placed themselves at the middle score of 3. Furthermore, 21.1% of people chose 4 or 5 when in practice 8.8% of people are actually popular according to the criteria we set.

There could be multiple reasons for this. Firstly, there is the criteria that we set (popular people are the one's who have above 25 friends). It is not necessary, that individuals who participated in the survey, necessarily have the same criteria. They could have a more relaxed criteria for popularity which explains the overestimation in the results.

Secondly, it could be that they did estimate correctly but a lot of their friends either did not choose them back or did not take part in the survey which skewed our results.

Lastly, it could be the Lake Wobegon effect at play. When it comes to desirable characteristics, people tend to illude their perception and oversell themselves. It could be that due to this most people who are less popular placed themselves in the mildly popular (score of 3) category and some of them who were actually mildly popular placed themselves in highly popular category (score of 4 and 5).

Similarly, the results of figure 8 are also skewed to the right. As the ability to make friends is also a desirable trait, it is very much likely that the Lake Wobegon effect is at work here too. Least people think they have difficulty making friends (26.9%). Many people think they are good at making friends (36.5%). Almost the same fraction also think they are highly good at making friends (35.5%). As in the previous question, most people probably overestimate themselves due to Lake Wobegon and hence we get these results.

This shows that class size paradox and Lake Wobegon effect do not go hand in hand and often are contrary to each other.

4.8 Attention in Friendship

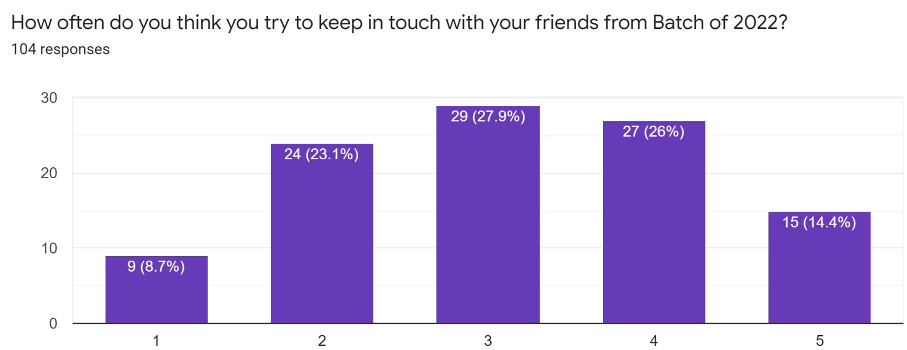

Fig. 9. Attention an individual thinks they give

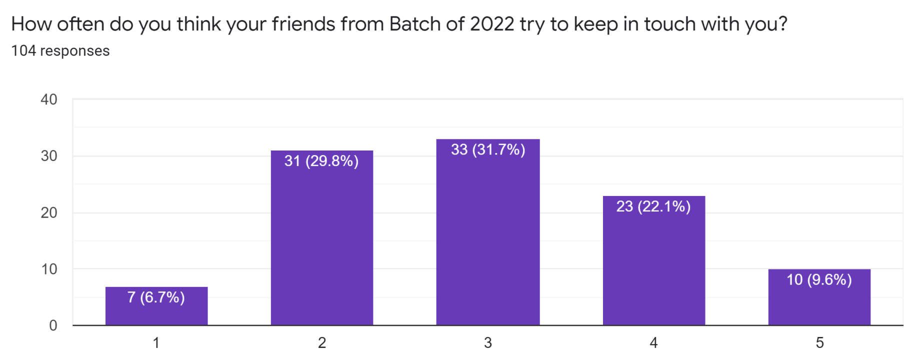

Fig. 10. Attention an individual thinks they get

As we saw earlier, an individual with less friends can give more attention to their friends while one with many friends would give less to them.

In figure 9, we tried to capture the amount of attention that people think they give to others. Speaking from the perspective of attention economics, there should have been more people choosing 4 and 5 as most people have less friends and hence would give more attention. This is true in the results obtained. However, a good fraction of people (27.9%) also chose 3.

Figure 10 showed the attention that an individual thinks they get. Once more according to attention economics, most people should have chosen 1 and 2 as there are many people with less friends and hence they would get less attention from their popular friends according to friendship paradox. This is true in the results obtained (39 people chose 1 and 2). However, as in the previous case, there are a good amount that chose 3 (31.7%).

Some features of the results obtained in these two questions do align with the idea of attention economics but there are some discrepancies. These could be mostly because of smaller sample size, less variance in the type of people and lack of participation from almost 40% of the batch of 2022. An interesting insight would also be to consider the effect of inverse of Lake Wobegon effect where people mostly think of themselves as below average and place themselves at the score of 3 in the case of figure 9. However, it is less likely compared to the other reasons.

4.9 One-sided Tale

Each individual did choose some people who they thought were their friends but their 'friends' did not think so. We found the weighted average of such one-sided friendships. There were a total of 982 choices of people that were one-sided. This means that on average, 9.63 friendships were one-sided for a single individual. These could be a likely factor for some of the biases in the perspectives of people and due to which the results that we obtained were off from the theory at some places.

4.10 Conclusion

Friendship is a two-sided relation - where each individual has different perceptions attached to it. Often one side feels advantaged and the other feels deprived. This idea of friendship paradox that most people feel deprived was prevalent in the context of Habib University's batch of 2022 as well. The structure of the network portrayed the reason for this through a power law adhering degree distribution.

However, there were probable traces of a phenomena like Lake Wobegon effect in the results. There were also some results that contradicted with the idea of class size paradox. Although, there is a possibility that we can not sufficiently determine whether the sample data collected was large and variable enough to study the effect of such phenomenon.

Friendship, definitely, is more of a subjective matter than an objective one. The experiences, perceptions, illusions and psychological phenomenon that affect one's thought process regarding friendships are complex and manifold. One thing is sure, there are two stories to every single 'friendship'.

References

1. Feld, S. L. (1991). Why Your Friends Have More Friends Than You Do. American Journal of Sociology, 96(6), 1464-1477. https://doi.org/:10.1086/229693
2. Coleman, James S. (1961). The Adolescent Society. New York: Free Press
3. Aparicio, S., Villazón-Terrazas, J., Álvarez, G. (2015). A Model for Scale-Free Networks:Application to Twitter. Entropy, 17(12), 5848-5867. https://doi.org/:10.3390/e17085848
4. David Hemenway (1982) Why Your Classes Are Larger than "Average", Mathematics Magazine, 55:3, 162-164, https://doi.org/: 10.1080/0025570X.1982.11976974
5. Maxwell, N. L. (1994). The Lake Wobegong effect in student self-reported data. The American economic review, 84(2), .
6. Yu, S., & Kak, S. (2013). Social Network Dynamics: An Attention Economics Perspective. Social Networks: A Framework of Computational Intelligence Studies in Computational Intelligence, 225-258. doi:10.1007/978-3-319-02993-1_11
7. Serafino, M., Cimini, G., Maritan, A., Rinaldo, A., Suweis, S., Banavar, J. R., & Caldarelli, G. (2021, January 12). True scale-free networks hidden by finite size effects. Retrieved from https://www.pnas.org/content/118/2/e2013825118

5 Appendix

[1]

5.1 Appendix A

The mean number of friends' friends is just the total number of friends' friends divided by the number of friends. To determine the total number of friends' friends, consider that each individual is a friend x_i times and has x_i friends, so that individual contributes x_i friends' friends x_i times, a total of ${x_i}^2$ friends' friends. Thus, the total number of friends' friends (the numerator) is simply this quantity summed over all individuals, $\sum {x_i}^2$. The total number of friends (the denominator) is simply the number of friends of each individual, x_i, summed over all individuals, $\sum x_i$. Thus, the mean number of friends' friends is just $= \frac{\sum {x_i}^2}{\sum x_i} =$. Some relatively simple algebra shows that this can be expressed as a function of the mean $(\frac{\sum x_i}{n})$ and the variance $(\frac{\sum {x_i}^2}{n} - mean^2)$. Thus, the mean number of friends' friends is: $= \frac{\sum {x_i}^2}{\sum x_i} = =$ mean(x) + variance(x)/mean (x).

[1] $https://github.com/AliSyedHammad/Networks_FinalReport$

5.2 Appendix B

To prove that a scale free network fulfills the small world property, the numerical of degree distribution, clustering coefficient and the average path length were computed for 50 million users.The degree distribution of the Twitter network was seen to follow power law - which means there were few users with very large number of followers and many users with very few friends and followers. The clustering coefficient of a twitter network was computed to be 0.096 (96%), which suggests that Twitter users tend to form clusters according to friendship links.Since the sample size was large, a heuristic method - using Breadth first search and a series of steps to find the path via the hub sub-network - to calculate the average path length was also introduced. The maximum path length in the hub sub-network was found to be three. The numerical confirm that Twitter, is a scale-free, small number with a high clustering coefficient.

5.3 Appendix C

For a population of $\boldsymbol{M}$ individuals divided into $\boldsymbol{N}$ groups, and let X_i denote the size of the ith group, $1 \leq i \leq N$. The expected number of people in a randomly selected group ("average class size") can be calculated as:

$$\bar{X} = \frac{\sum X_i}{N} = \frac{M}{N}$$

The expected size of a group containing a randomly selected individual is given by:

$$X* = \sum(\frac{X_i}{M})X_i = \frac{\sum (X_i)^2}{M}$$

The difference between the average class size and the expected size is found to be directly proportional to the variance in the sizes of groups and inversely proportional to the average group size.