The Friendship Paradox:
Why your friends being more popular than you could predict the future

A peculiar collision of maths and sociology was observed by Scott L. Fled, expressed in detail in his 1991 publication of "Why Your Friends Have More Friends Than You Do" (Feld, 1991). The publication presents "The Friendship Paradox"; put simply, your average number of friends, compared to the average number of friends of your friends, is lower. How does this make sense? Especially given that friendships are bilateral: if x is friends with y then y is friends with x, so why, in a social network, this "paradox" true? This seemingly paradoxical argument can be put down to a consequence of sampling bias, and when you consider that you are more likely to be friends with someone who is popular (who has more friends than yourself) versus somebody who has very few friends, you can begin to grasp the concept more easily.

Consequences of understanding this paradox have very significant applications when it comes to our understanding of society today. By realising that there are people who act as central and social "hubs" in our society, scientists can monitor spread of trends, whether this is spread of disease, behaviour, social norms or information. Furthermore, by understanding this paradox on an individual scale, it can have beneficial psychological impacts and alter our personal view of reality.

Sampling bias

The paradox is a form of sampling bias as people with more friends are more likely to be in a friend group; specifically, "**self-selection bias**". This occurs when individuals select themselves into the group in situations where certain characteristics of people cause them to appear more significantly, creating abnormal conditions. (Contributors, 2020).

In the friendship paradox, individuals commit self-selection bias by becoming friends with individuals who are more popular; the characteristic of having a friend with relatively higher number of friends appears more frequently. Hence if a person is always friends with a more popular person, their average number of friends is always lower than the average number of friends of their popular friend.

Variance

In this paradox, it is useful for us to understand **variance** and what is represents. Variance (σ^2) is the average of squared difference from the mean. It weights data points further from the mean more heavily. Relating this to the friendship paradox, anomalies further from the average are people with a higher number of friends, therefore appearing in more friendships of other people.

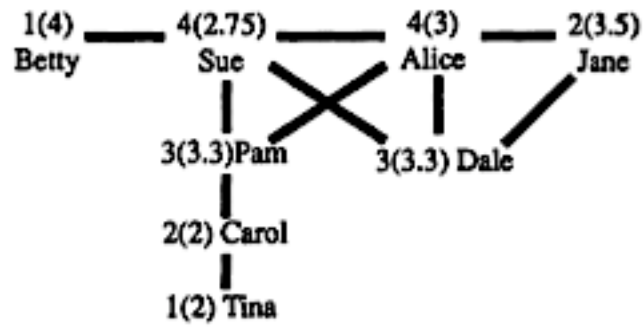
The general formula is given by the equation:

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

The Original Experiment

In the experiment referenced by Feld, *The Adolescent Society,* (Coleman, 1961)*,* data was collected across 12 high schools; where individuals were asked to name their friends. “Friend”, in this study, is a symmetric relationship, therefore only pairs of students who named one another were considered.

Feld considered the friendship group of 8 girls, illustrated below:



The number beside each name is her number of friends. The number in parentheses beside each name is the mean number of friends of her friends.

Name	No. friends ($\Sigma\gamma_i$)	No. friends of friends ($\Sigma\gamma_j$)	Mean number of friends of her friends ($\Sigma\gamma_j/\gamma_i$)
Betty	1	4	4
Sue	4	11	2.75
Alice	4	12	3
Jane	2	7	3.5
Pam	3	10	3.33
Dale	3	10	3.33
Carol	2	4	2
Tina	1	2	2
Total	20	60	23.92
Mean average	2.5	3*	2.99

*This average was calculated using

$$\frac{\Sigma\gamma j}{\Sigma\gamma i}$$

As the girls have a total of 20 friends within the group.

The average number of individual's friends is 2.5, but the average number of friend's friends is 3.

The average given in the second column is not the average number of friendships that the friend of an average individual has, it's the average number of friends of friends when considering number of friendships within the overall group (20).

The 3rd column's average is the mean number of friends of friends of a mean individual.

The two distributions to consider: the distribution of friends of an individual and the distribution of friends' friends. The latter is a weighted version of the former, with girls who have more friends weighted heavily, as they appear in friendships as many times as they have friends. This, therefore, means the distribution of the number of friend's friends always has a higher average than the average of the original distribution.

Relationship between the averages

As seen, the average for the original distribution is:

$$\frac{\Sigma\gamma_i}{n}$$

Where n = number of individuals.

The average number of friends of friends is:

$$\frac{\Sigma\gamma_j}{\Sigma\gamma_i}$$

With $\Sigma\gamma_j$ = the total number of friends of friends.

can also be written as:

$$\Sigma\gamma_j$$

$$\Sigma\gamma^2{}_i$$

This is because, for the total number of friends of friends, each person is counted as many times as they have friends.

From this the mean number of friends of friends is written as:

$$\frac{\Sigma\gamma^2{}_i}{\Sigma\gamma_i}$$

The mean individual's mean number of friends is just the sum of $j's$ over their corresponding $i's$ divided by the total number of individuals, as seen in the table:

$$\frac{\Sigma\left(\frac{\gamma_j}{\gamma_i}\right)}{n}$$

Incorporating variance and its effect

The mean number of friends of friends can be written as a function of the mean and the variance:

$$\frac{\Sigma\gamma^2{}_i}{\Sigma\gamma_i} = \left(\frac{\Sigma\gamma_i}{n}\right) + \sigma^2 \div \left(\frac{\Sigma\gamma_i}{n}\right)$$

$$= \text{mean}(\gamma) + \left(\frac{\text{variance}(\gamma)}{\text{mean}(\gamma)}\right)$$

To show this, take an image of a wheel composed of *n* individuals and *n-1* number of spokes, where the person in the centre has *n-1* friends and the remaining *n-1* friends only have 1 friend, the centre person.

$$mean\ (\mu) = \frac{2(n-1)}{n}$$

$$variance\ (\sigma^2) = \left(\left((n-1) - \frac{2(n-1)}{n}\right)^2 + \left(1 - \frac{2(n-1)}{n}\right)^2 \times (n-1)\right) \times \left(\frac{1}{n}\right)$$

$$= \left(\left(\frac{(n-1)(n-2)}{n}\right)^2 + \left(\frac{2-n}{n}\right)^2 \times (n-1)\right) \times \left(\frac{1}{n}\right)$$

$$= \left(\frac{(n-1)^2(n-2)^2 + (n-2)^2(n-1)}{n^2}\right) \times \left(\frac{1}{n}\right)$$

$$= \frac{n(n-2)^2(n-1)}{n^3}$$

$$\sigma^2 = \frac{(n-2)^2(n-1)}{n^2}$$

In terms of n:

$$\frac{\Sigma\gamma^2_i}{\Sigma\gamma_i} = \frac{(n-1)^2 + 1^2 \times (n-1)}{2(n-1)}$$

$$= \frac{n^*}{2}$$

*please note that this simplification is only true for a wheel with n number of individuals and n-1 spokes. It has been fully simplified in order to show the function of variance and the mean.

And substituting the mean and variance into the equation and evaluating:

$$\left(\frac{\Sigma\gamma_i}{n}\right) + \sigma^2 \div \left(\frac{\Sigma\gamma_i}{n}\right) = \frac{2(n-1)}{n} + \frac{(n-2)^2(n-1)}{n^2} \div \frac{2(n-1)}{n}$$

$$= \frac{2(n-1)}{n} + \frac{(n-2)^2}{2n}$$

$$= \frac{4(n-1) + (n-2)^2}{2n}$$

$$= \frac{4n - 4 + n^2 - 4n + 4}{2n}$$

$$= \frac{n^2}{2n}$$

$$= \frac{n}{2} = \frac{\Sigma\gamma^2_i}{\Sigma\gamma_i}$$

We see that it is true.

Therefore, the mean number of friends of friends,

$$= \text{mean}(\gamma) + \left(\frac{\text{variance}(\gamma)}{\text{mean}(\gamma)}\right)$$

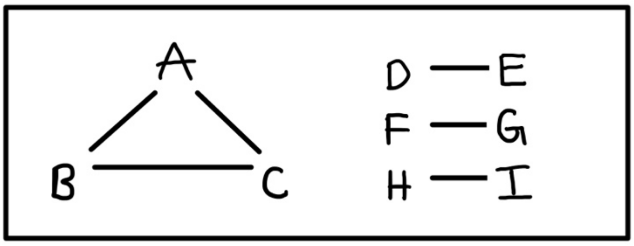
is always at least as great as the mean of the individuals. We can also see that this average increases with variance. In real life terms, if your friends all have extremely differing numbers of friends, you are more likely to find that your friends, on average, are significantly more popular than yourself.

Effect of correlation

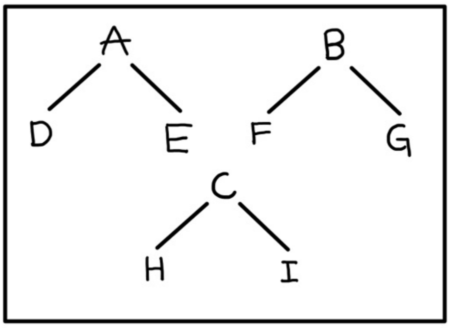
Another factor affecting the mean individual's mean number of friends is correlation, the arrangement of friendships in a group. The final two step averaging process to find the mean individuals mean number of friends weights each friend. The average number of friends' friends differs to that of the average individual. Taking the first distribution of the girls at the high school, Betty's average, 4, based on her only friend Sue, who is the most popular, is weighted equally to Sue's average of her 4 friends, 2.75. The value of Betty's average number of friends' friends is more than that of Sue, and as Betty's one friend is popular, the final average is higher than the average number of friends.

There is possibility that some people in a group are friends with friends who have similar friendship volumes; below is an example of how correlation impacts the final average and is a result of how the final two step averaging process to find the mean individuals mean number of friends weights each friend accordingly, with the first example a perfect positive correlation, and the second, a perfect negative correlation.

To demonstrate, see 2 diagrams, with individuals represented as letters A to I. Each yields the same average number of friends per person, as the number of people, and each letter's number of friends, is constant:



A. Perfect Positive Correlation



B. Perfect Negative Correlation

The average number of friends in both of the above diagrams is 1.33. However, the value of an average individual's mean number of friends of friends' changes.

In diagram A,

$$\Sigma\left(\frac{\gamma_i}{\gamma_i}\right) = 12$$

And therefore, the mean individual's mean number of friends is:

$$\frac{12}{9} = 1.33$$

as the correlation is perfect and positive, the minimum value of the weighted mean is reached; achieved when all of the friends have the same number of friends and their friends, resulting in the weighted mean just being the mean number of friends for an individual.

However, this is not true for diagram B:

$$\Sigma\left(\frac{\gamma_i}{\gamma_i}\right) = 15$$

And the mean individual's mean number of friends is:

$$\frac{15}{9} = 1.67$$

In diagram B, pairs of friends that A, B, C were friends with were not related, having a negative correlation. As negativity of the correlation increases amongst the group, the greater the mean individuals' mean number of friends' friends.

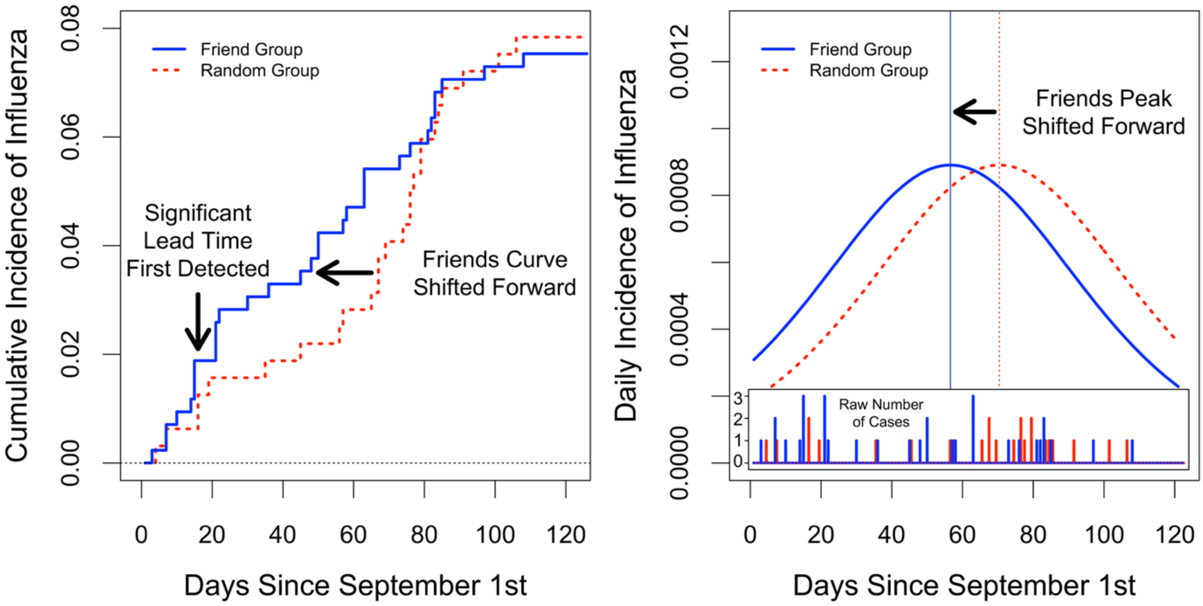
Real world applications

This paradox ultimately suggests that, as friendship groups act like a social network, within social networks there are always people who are the most popular, who have contact with the greatest number of people, acting as "social hubs". Through understanding the structure of social networks, scientists would be able to predict the outbreak of disease, predict fluctuation within the economy, and more efficiently distribute vaccines.

Harvard flu outbreak

In 2009, the paradox was used to overcome the difficulties that are usually faced when trying to monitor a social network, those being how time consuming and expensive this usually is. James Fowler and Nicholas Christakis, both professors at Harvard University, carried out and experiment used this paradox to monitor the 2009 flu epidemic among a network of 744 students (Christakis & Fowler, 2010).

They contacted 319 university undergraduates who gave names of 425 friends. The theory was that, as one of the 319 graduates is more likely to be less popular than the friend that he or she named, the named group of 425 students would develop the flu faster than the random sample. In theory, they are in contact with relatively more people. They followed the 2 groups and found that, on average, the group of 425 students manifested the flu 16 days prior to the random group.



The results of this experiment could be used on a larger scale to track and potentially reduce the spread of an outbreak,

> *"Public health officials track epidemics by following random samples of people or monitoring people after they get sick. That approach only provides a snapshot of what's currently happening. By simply asking members of the random group to name friends, and then tracking and comparing both groups, we can predict epidemics before they strike the population at large. This would allow an earlier, more vigorous, and more effective response." – Nicholas Christakis*

Further applications

In a subsequent TED talk given by Christakis (Christakis, 2010) he suggested that by using this information to structure a social network, the prevention of a disease through the distribution of vaccinations could be made more efficient. He referred to the term "herd immunity" which occurs when 96% of the population is vaccinated. Traditionally, to do this, as is being done currently, the distribution of the vaccine is based on who is the most vulnerable to the disease itself. For example, COVID-19 vaccines and their distribution according to a list of descending age groups. However, scientists have suggested that, given a random sample of 1000 people, if you take 300, ask them to nominate a friend each to be given the vaccine, it is the equivalent of vaccinating 960 of the population, due to the structure of social networks. In other words, by using the friendship paradox, 30% becomes the new 96%. This method could be implemented if the supply of vaccines was limited; not only would time efficiency increase, but also cost efficiency.

Not only could this be used to monitor the spread of pathogens, but also used to predict behaviours which may have an economic impact. For example, through passive observation, the network of truckers and their purchases of fuel. By understanding the social network, if there is a "blip up" in their purchases of fuel, it could indicate that a recession is about to end.

To conclude, consider this on an individual scale; understanding the friendship paradox could be beneficial to our mental health. When addressing imposter syndrome: the internal experience of believing that you are not as competent as others perceive you to be (Cunic, 2021), the understanding that it is almost predetermined that the people around you are doing more of the thing you are doing, could be reassuring. For example, writing, if you're part of a writer's community and you feel that others are writing more than you, or studying as a student, who believes they are not working as hard as their fellow classmates for their university application. Chances are, you have surrounded yourself with people whom you aspire to be like or have similar interests to, and the friendship paradox suggests that it this observation is expected.

Your friends are more popular than you, but it's a good thing.

By Rachael McEvoy

Bibliography

Christakis, N., 2010. *How social networks predict epidemics.* [Online]
Available at:
https://www.ted.com/talks/nicholas_christakis_how_social_networks_predict_epidemics/transcript?language=en
[Accessed March 2021].
Christakis, N. & Fowler, J., 2010. *Social Network Sensors for Early Detection of Contagious Outbreaks.* [Online]
Available at: https://doi.org/10.1371/journal.pone.0012948
[Accessed 21 March 2021].
Coleman, J. S., 1961. *The Adolescent society : the social life of the teenager and its impact on education.* [Online]
Available at: https://www.semanticscholar.org/paper/The-Adolescent-society-%3A-the-social-life-of-the-and-Coleman/7da25503d503507ab201ed641e3ef4e920cfa90b#paper-header
[Accessed 23 March 2021].
Contributors, W., 2020. *Self-selection bias.* [Online]
Available at: https://en.wikipedia.org/wiki/Self-selection_bias
[Accessed 23 March 2021].
Contributors, W., 2021. *Friendship Paradox.* [Online]
Available at:
https://en.wikipedia.org/w/index.php?title=Friendship_paradox&oldid=1003471368
[Accessed 23 March 2021].
Cunic, A., 2021. *What is Imposter Syndrome?.* [Online]
Available at: https://www.verywellmind.com/imposter-syndrome-and-social-anxiety-

disorder-
4156469#:~:text=Impostor%20syndrome%20(IS)%20refers%20to,perfectionism%20and%20the%20social%20context.
[Accessed 25 March 2021].
Feld, S. L., 1991. “Why Your Friends Have More Friends Than You Do.”. *American Journal of Sociology,* 96(6), pp. 1464-1477.