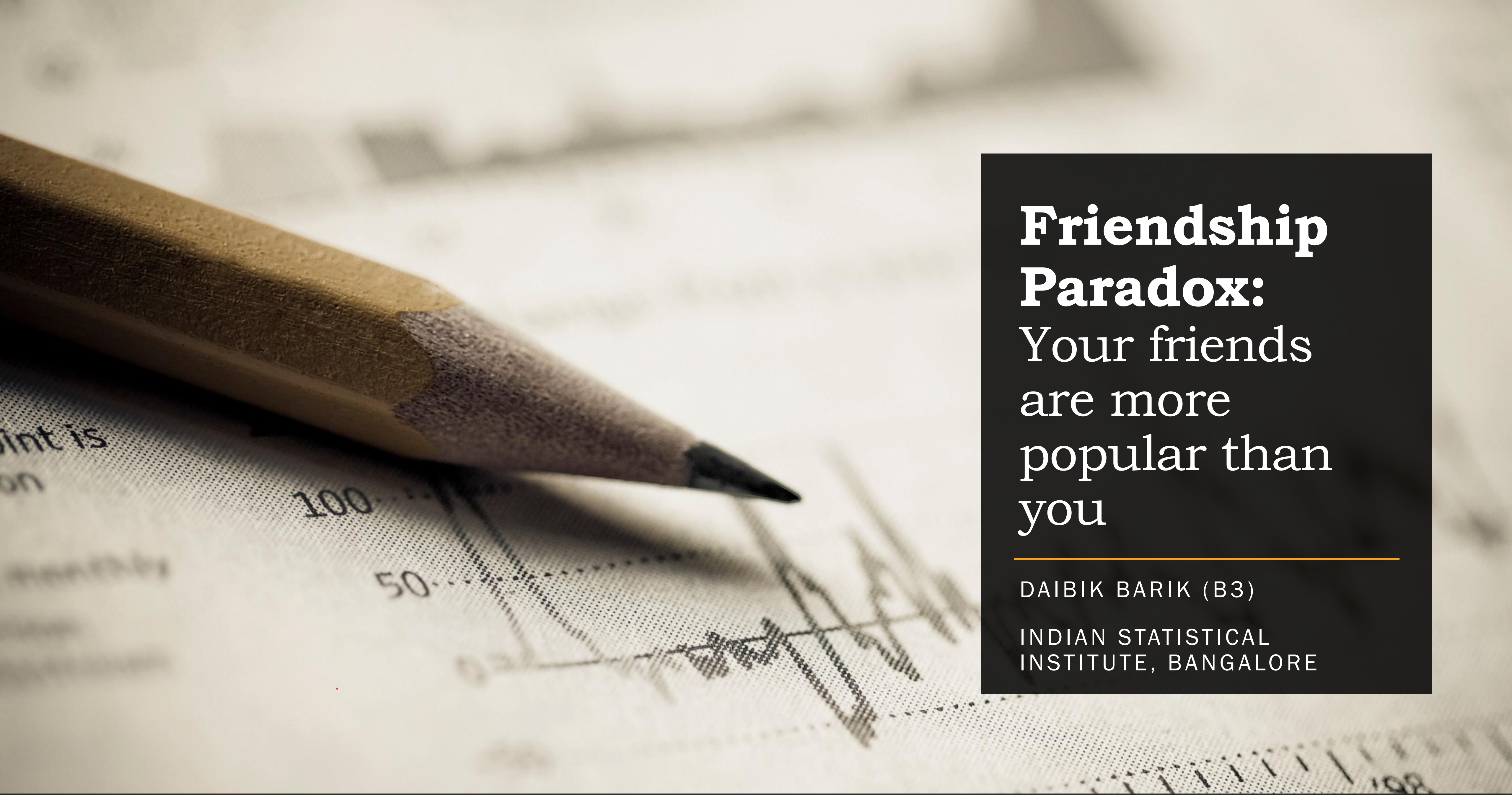



# **Friendship Paradox:** Your friends are more popular than you

DAIBIK BARIK (B3)

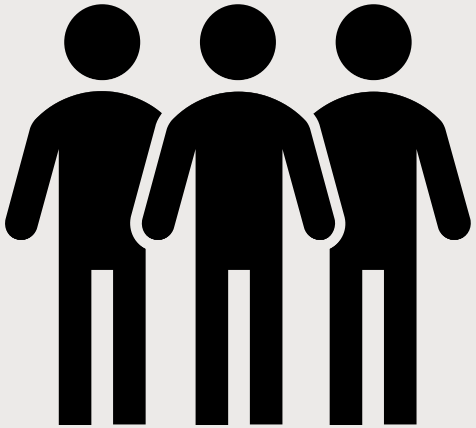
INDIAN STATISTICAL INSTITUTE, BANGALORE

# Ever feel like everyone's more popular than you?

So there may be times when you feel alone, probably because your friends are talking with their other friends that doesn't include you?

Let me tell you, everyone feels that in some point of their lives. Probably because of the false-lens of social media or because you're being too judgemental on yourself, whatever it is, it's not a big deal.

That's *probably* not because you're not likable, and I'm here to make you feel a bit better by saying it's…. Absolutely TRUE !

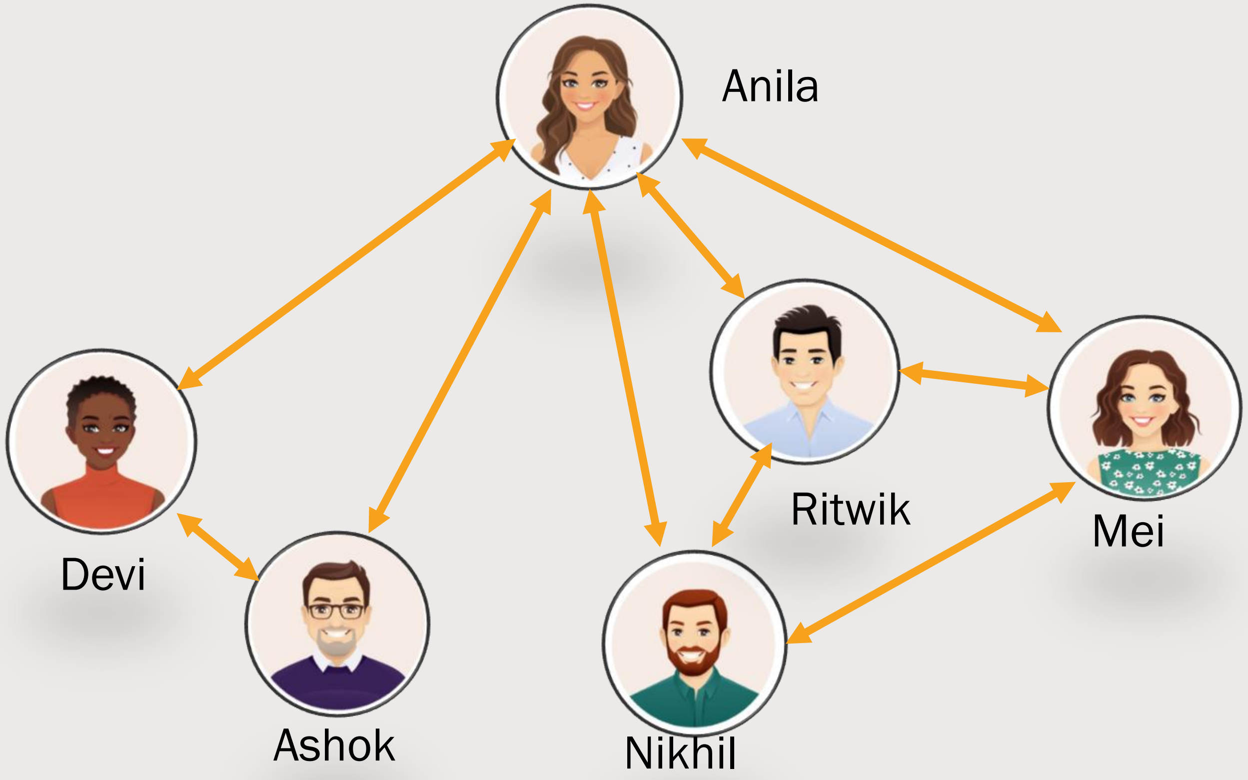


# What? Really? How?

See I don’t want to make you feel bad or just give you a psychology lesson on why you shouldn’t feel that way, but to show it, or rather *mathematically prove* it’s existence.

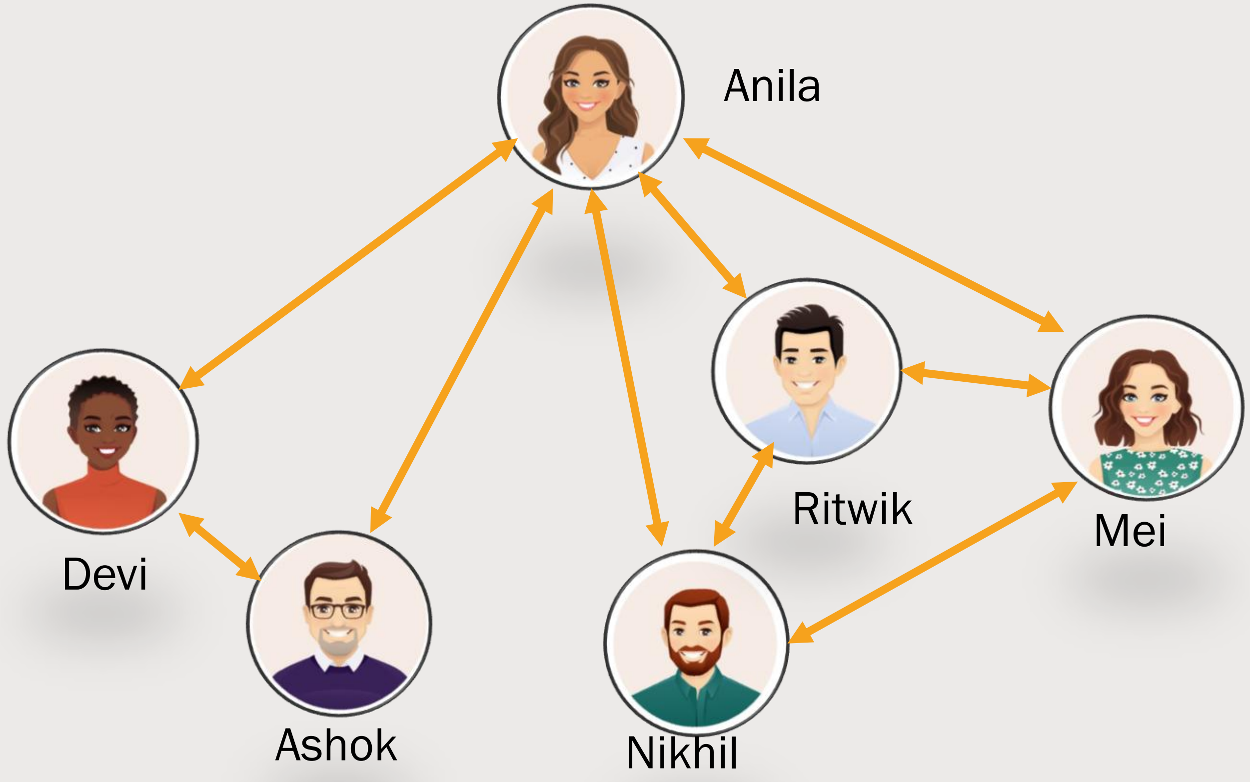
This is what’s called the **Friendship paradox**, which says that *on average, an individual's friends have more friends than that individual*. I will first explain the model with an example and leave it for you to find the truth behind it.

# **An Example:** A Friend Network




We Calculate the number of friends of each person and take their average.

# **An Example:** Listing out Friends




| Name | Number of friends |
|---|---|
| Anila | 5 |
| Devi | 2 |
| Ashok | 2 |
| Nikhil | 3 |
| Ritwik | 3 |
| Mei | 3 |
| **Total :** | **18** |

# **An Example:** Calculating average number of friends

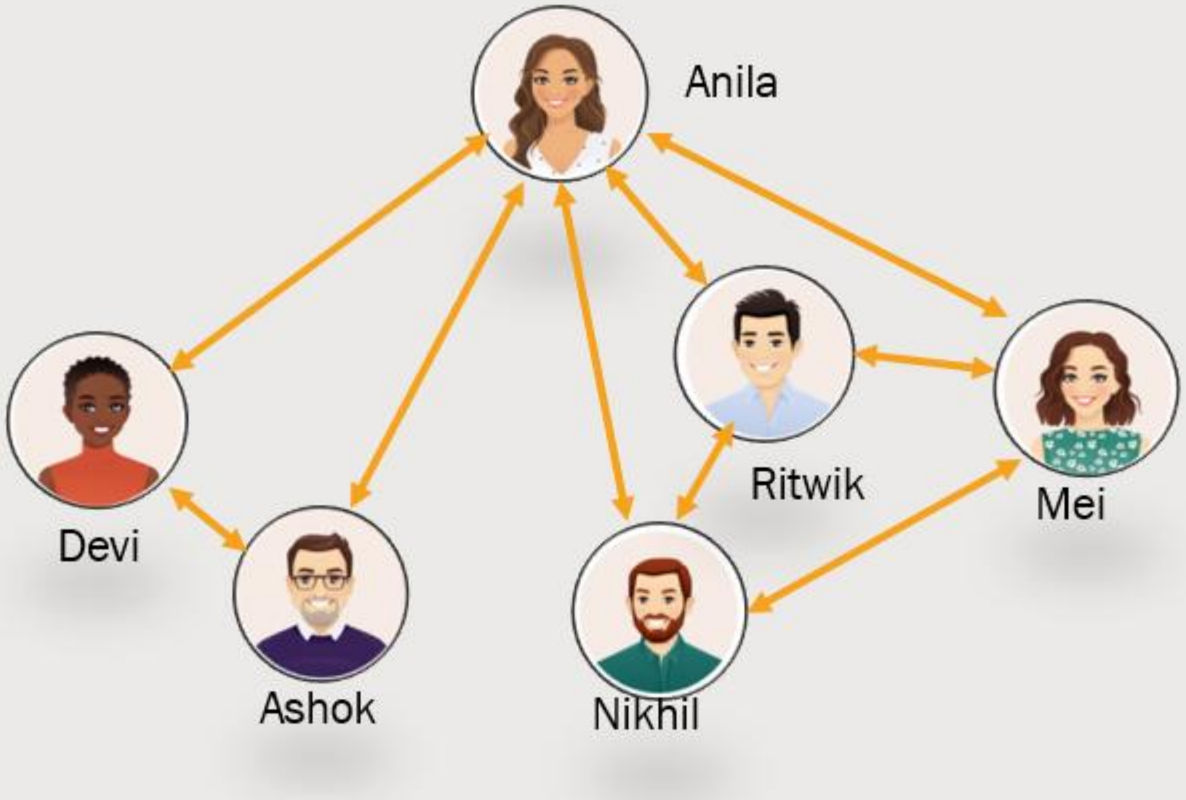
| Name | Number of friends |
|---|---|
| Anila | 5 |
| Devi | 2 |
| Ashok | 2 |
| Nikhil | 3 |
| Ritwik | 3 |
| Mei | 3 |
| **Total :** | **18** |

Average number of friends

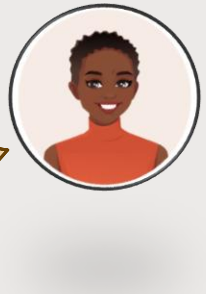
$$= \frac{\text{Total number of friends}}{\text{Number of people having friends}}$$

$$= \frac{18}{6} = 3$$

# **An Example:** Listing out Friends of friends

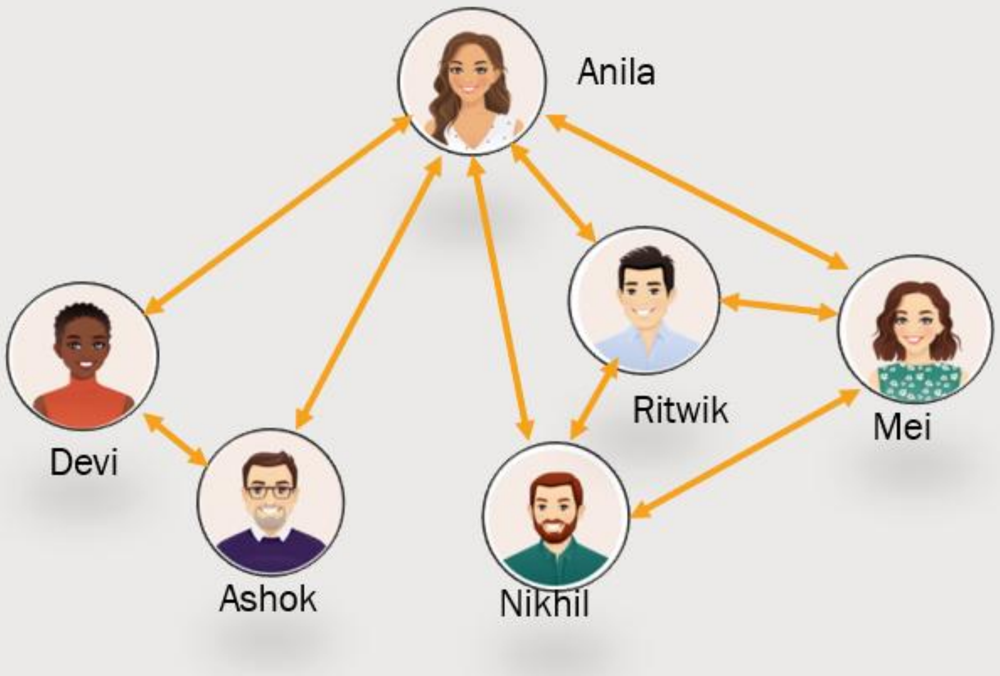



*Devi: I have 2 friends, Anila and Ashok*



| | |
|---|---|
| Anila | 5 |
| Ashok | 2 |
| **Total:** | **7** |

# An Example: Averaging out Friends of friends




| Names | Friends (No. of friends) | Average |
|---|---|---|
| Anila (5) | Devi (2) + Ashok (2) + Nikhil (3) + Ritwik (3) + Mei (3) | 2.6 |
| Devi (2) | Anila (5) + Ashok (2) | 3.5 |
| Ashok (2) | Anila (5) + Devi (2) | 3.5 |
| Nikhil (3) | Anila (5) + Ritwik (3) + Mei (3) | 3.6 |
| Ritwik (3) | Anila (5) + Ritwik (3) + Mei (3) | 3.6 |
| Mei (3) | Anila (5) + Ritwik (3) + Mei (3) | 3.6 |
| **Total Average (initially 3):** | | 4.08 |

# **An Example:** Averaging out Friends of friends

| Names | Friends (No. of friends) | Average |
|---|---|---|
| Anila (5) | Devi (2) + Ashok (2) + Nikhil (3) + Ritwik (3) + Mei (3) | 2.6 |
| Devi (2) | Anila (5) + Ashok (2) | 3.5 |
| Ashok (2) | Anila (5) + Devi (2) | 3.5 |
| Nikhil (3) | Anila (5) + Ritwik (3) + Mei (3) | 3.6 |
| Ritwik (3) | Anila (5) + Nikhil (3) + Mei (3) | 3.6 |
| Mei (3) | Anila (5) + Ritwik (3) + Nikhil (3) | 3.6 |
| **Total Average (initially 3):** | | 4.08 |

Average Number of friends of Friends

$$= \frac{\text{Total Number of friends of all friends}}{\text{Total Number of friends}}$$

# **An Example:** Calculation

$$= \frac{(2+2+3+3+3)+(5+2)+(5+2)+(5+3+3)+(5+3+3)+(5+3+3)}{18}$$

$$= \frac{5\times5+3(3\times3)+2(2\times2)}{18} = \frac{60}{18} \approx 3.34 > 3 \text{ (Prev. obtained Avg.)}$$

Anila

Ritwik, Nikhil, Mei

Devi, Ashok

Total Number of friends of everyone

# This was just a show..

Yeah but we learnt a few things from it right?

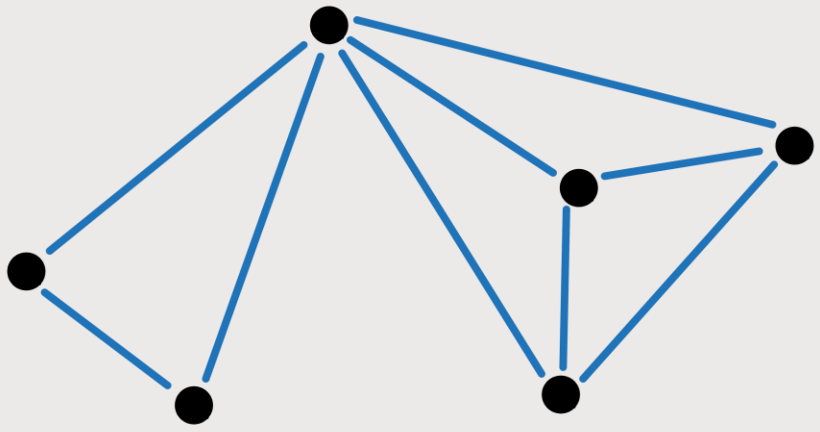
1) Like how the **popular friends** (for eg. Anila in this case) **contribute more** to the weighted average.

2) Also <u>**NOTE:**</u> This paradox speaks **only for the average population** and is **not true for every person**. For eg: Take the person with most number of friends (exists!, say $n$), and when taken his/her friends avg. number of friends it has to be less than $n$ (since it's the maximum).

Now its our turn to prove it mathematically.

# **Mathematical formulation:** Modelling out the Problem

To tackle this model, we use the concept of **simple undirected graphs.**

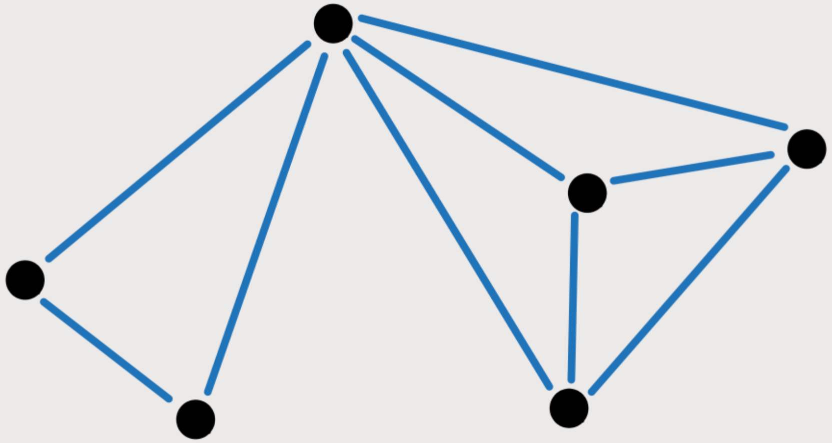


$G = (V, E)$

Consider every friend as a node (or vertex). Now we assume friendships are mutual, i.e., if $A$ is a friend of $B$ then $B$ is also a friend of $A$. So we can represent it as an undirected edge between $A$ and $B$.

Hence we have the network as a Graph $G = (V, E)$, where $V$ is the set of all friends and $E$ is the set representing the friendships between them.

# **Mathematical formulation:** Average #Friends ($\mu$)



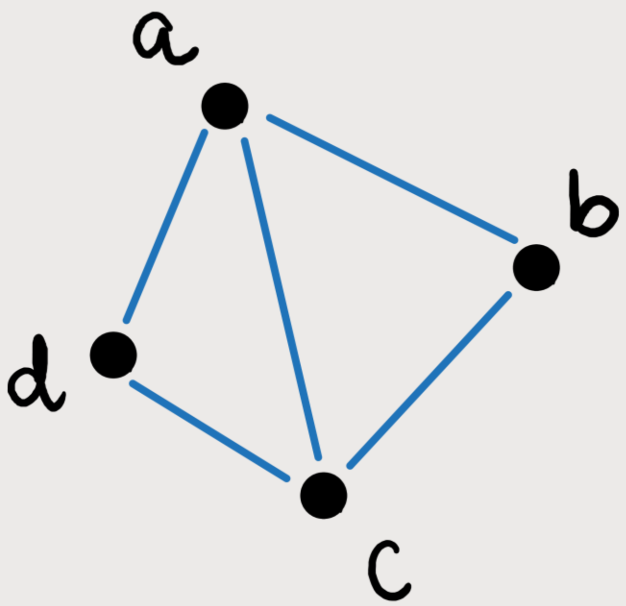
$G = (V, E)$

Firstly, let's find the *average number of friends.* We call the it the **mean** ($\mu$).

Ideally, we go to every friend and count their number of friends, i.e. here we take the summation over the degrees of every vertex.

Mathematically, $\boxed{\mu = \frac{\sum_{v \in V} d(v)}{|V|}}$

# <u>**Claim:**</u> $\sum_{v \in V} \mathrm{d}(v) = 2|E|$




$G = (\{\mathrm{a}, \mathrm{b}, \mathrm{c}, \mathrm{d}\}, \{aa, ab, \dots, cd\})$

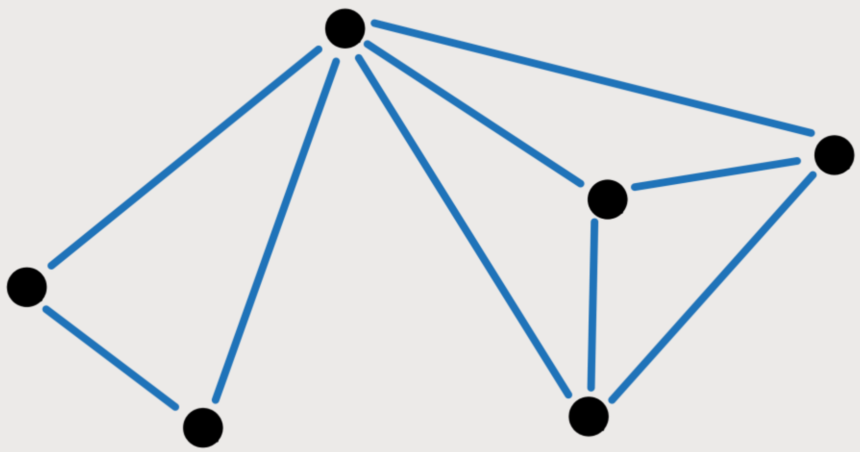
Take any general graph $G$, name the edges according to the conventions.

Now, while counting the degree of say *vertex* $a$, we take the *edge* $ac$ into consideration once. We take the *edge* $ac$ into consideration again while counting the degree of *vertex* $c$.

Drawing an injective map from degrees of every vertex and the edges in the graph, we find it maps to every edge **exactly twice.**

Hence $\sum_{v \in V} \mathrm{d}(v) = 2|E|$

# **Mathematical formulation:**
# Average #Friends of friends $(\mu_f)$



$$G = (V, E)$$

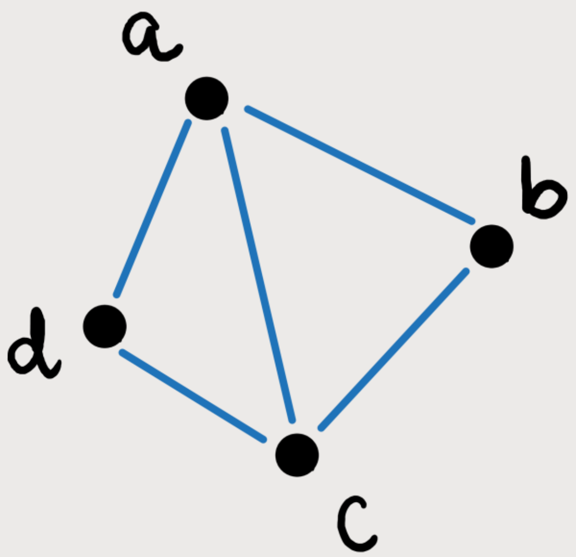
Secondly, let's find the *average number of friends of the friends.* We call the it the **friends' mean** $(\mu_f)$.

Ideally, we go to every friend, then to every friend of that person and count their number of friends, i.e. here we take the double summation over the degrees of the neighbouring vertices of every vertex.

Mathematically, $\boxed{\mu_f = \frac{\sum_{v \in V} \sum_{u \in N(v)*} d(u)}{\sum_{v \in V} d(v)}}$

$*N(v)$ means the neighbourhood of $v$, i.e. $\{u: uv \in E\}$

# Claim: $\sum_{v \in V} \sum_{u \in N(v)} d(u) = \sum_{v \in V} d(v)^2$




$G = (\{\mathrm{a, b, c, d}\}, \{aa, ab, \dots, cd\})$

Take any general graph $G$, name the edges according to the conventions.

Let's write out the LHS. We consider $a$'s friends (i.e. add $d(a)$) when we count for $d, b, c$ – who have $a$ as a friend. Therefore $d(a)$ is **summed** once for all friends of $a$, i.e. $d(a)$ times.

Since for all vertices $v$, we add $d(v)$ times $d(v)$; Hence

$$\sum_{v \in V} \sum_{u \in N(v)} d(u) = \sum_{v \in V} d(v)^2$$

# **Mathematical formulation:** Comparison

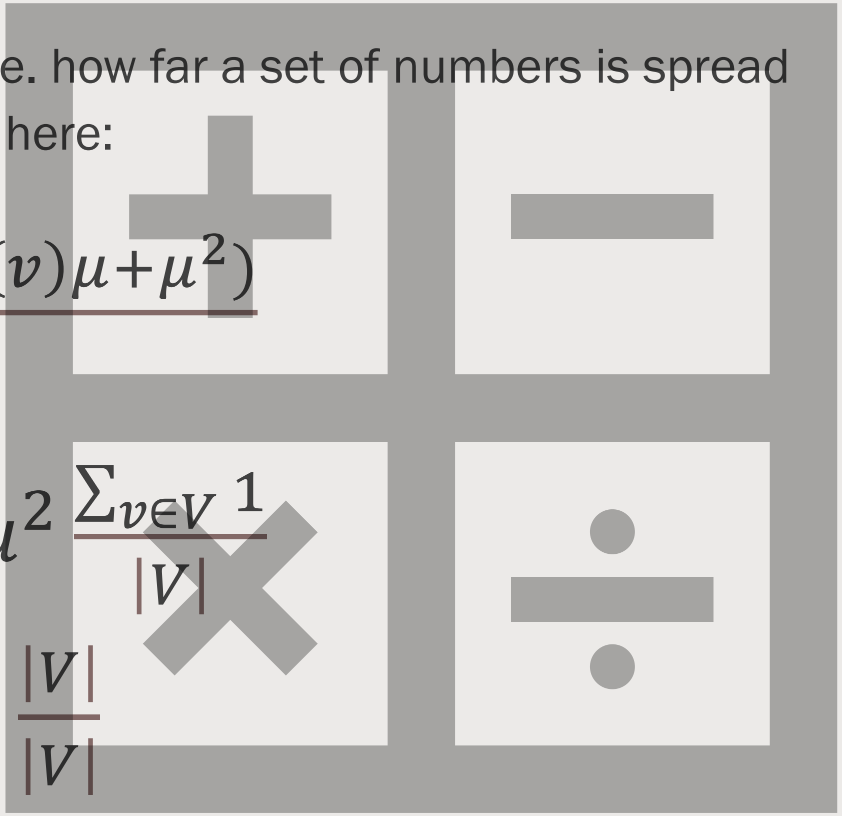
$$\mu = \frac{2|E|}{|V|}, \qquad \mu_f = \frac{\Sigma_{v \in V}\, d(v)^2}{2|E|}$$

Now, we have the 2 results, but we need to make some changes if we want to compare them.

<u>**Motivation:**</u> we notice the term $d(v)^2$ in the numerator of $\mu_f$, which makes us think, how to we manipulate to make the quadratic term disappear to that is it comparable to $\mu$?
And it clicks, we need to introduce the **Variance ($\boldsymbol{\sigma^2}$).**

# **Mathematical formulation:** Variance #Friends $(\sigma^2)$

**Variance** can be described as "*a measure of dispersion*", i.e. how far a set of numbers is spread out from their average value. We use the standard formula here:



$$\sigma^2 = \frac{\Sigma_{\nu \in V}(d(\nu) - \mu)^2}{|V|} = \frac{\Sigma_{\nu \in V}\,(d(v)^2 - 2d(v)\mu + \mu^2)}{|V|}$$

$$= \frac{\Sigma_{\nu \in v}\, d(\nu)^2}{|V|} - 2\mu \frac{\Sigma_{\nu \in V}\, d(\nu)}{|V|} + \mu^2 \frac{\sum_{\nu \in V} 1}{|V|}$$

$$= \frac{\Sigma_{\nu \in v}\, d(\nu)^2}{|V|} - 2\mu \times \mu + \mu^2 \times \frac{|V|}{|V|}$$

# Mathematical formulation:
Variance #Friends $(\sigma^2)$

$$\Rightarrow \sigma^2 = \frac{\Sigma_{\nu \in v}\, d(\nu)^2}{|V|} - 2\mu^2 + \mu^2 = \frac{\Sigma_{\nu \in v}\, d(\nu)^2}{|V|} - \mu^2$$

$$\Rightarrow \sigma^2 + \mu^2 = \frac{\Sigma_{\nu \in v}\, d(\nu)^2}{|V|}$$

$$\boxed{\Rightarrow (\sigma^2 + \mu^2)|V| = \Sigma_{\nu \in v}\, d(\nu)^2}$$

… and we substitute this value into the term $\mu_f$

# **Mathematical formulation:** Comparison by substitution

$$\therefore \mu_f = \frac{\Sigma_{v \in V}\, d(v)^2}{2|E|} = \frac{(\sigma^2 + \mu^2)|V|}{2|E|} = \frac{(\sigma^2 + \mu^2)}{\mu} = \mu + \frac{\sigma^2}{\mu}$$

$$\Rightarrow \mu_f = \mu + \frac{\sigma^2}{\mu} \geq \mu$$ (Hence Proved)

… phewww, that was too much at once.

# Understanding the Math.

It’s to say, kind of a **sampling bias** rather than a paradox, which is what we call an *error of statistical calculation.*

Similarly, its proven in various other cases, such as people’s romantic partners tend to have more romantic partners in average, people’s research collaborators tend to have more collaborators in average than them but strangest of all, people’s enemies have more enemies in average…

Let me give you an example of Sampling Bias to make it more understandable…

Frequently Asked Questions

# How large are classes?

Some introductory courses as well as several other popular courses have large enrollments. Yet, the median class size at Harvard is 12. Of the nearly 1,300 courses offered last fall, for example, more than 1,000 of them enrolled 20 or fewer students.

# **Sampling Bias:** An example

Take 5 Classes, say **A1 to A5,** with class sizes say 1 each for first four and 96 for the last one.

Average class size $= \frac{1+1+1+1+96}{5} = 20$

Even though the class mean is 20, we see about 96% of people are in a class size of 96!

This is because, weighted average class size $= \frac{4(1)+96(96)}{100} = 92.2$

Correcting Harvard's class size, their new average is 73.4 – which makes it an 232% increase!

→ DON'T BE FOOLED BY STATISTICS!!

# **Sampling Bias:** Methodology

This is not a new phenomenon, we always compare ourselves with people who are more advanced in the field of comparison, we say our comparison dataset is biased.

Say when we get to a premier institute, we still compare us with the people who top the exams or competition there, or say when we play sports, we compare ourselves to the people who are better than us – missing out on the ***randomness of the dataset.*** We leave out the people who are just learning the sport, for them we lie in the “comparable dataset”.

Now I’m not here to say if that’s wrong or right, but that’s what happens in the friendship paradox. We make friends with those who are more *“popular”, “likable”* or *“friendly”.* An introverted person, in general, might not have that many friends, or in this case, might not be friends with you. So your *“friends dataset”* is biased.


Q
A

# **An Application:** SuperSpreaders

In the recent past, during a flu outbreak, Scientists started tracking 2 groups of people.

One was a random set of people, and other was the group of their friends.

It was found that, the second group of people (being more popular), are more likely to **catch the flu early on, and then spread to more people**, due to **their high number of connections.**

The study went as expected, the '***Superspreaders***' caught the flu about 2 weeks (on average) earlier than the first group, thus giving us a method to catch disease outbreak and counter it faster.

# References

*Wikipedia contributors. (2025, March 12). **Friendship paradox.** Wikipedia.* *https://en.wikipedia.org/wiki/Friendship_paradox*

Zach Star. (2019, June 16). ***The Friendship Paradox | This is in your recommended because it relates to the spread of diseases*** [Video]. YouTube. https://www.youtube.com/watch?v=GEjhO65FYks

Giddon, K. (2021, December 13). **Why colleges' "average class size" statistics are misleading — and how to fix them**. *Medium*. https://kylegiddon.medium.com/why-colleges-average-class-size-statistics-are-misleading-and-how-to-fix-them-f00a59668695

*Feld, Scott L. (1991), "**Why your friends have more friends than you do**", American Journal of Sociology, **96** (6): 1464–1477, doi:10.1086/229693, JSTOR 2781907, S2CID 56043992.*


THANK YOU :)