The Friendship Paradox

Barum Park

Department of Sociology
Cornell University
b.park@cornell.edu

March 8, 2021

The Paradox

The friendship paradox states that, on average, your friends will have more friends that you have. It is called a paradox because most people believe that they have more friends than their friends (is this true?). Interestingly enough, Feld called this the "class size paradox," but his own terminology was not adopted in the literature. Formally, we might state the paradox as follows:

> Let $G(V, E)$ be a connected graph with $|V| = n$ nodes, $|E| = m$ edges, and a degree sequence $\text{Deg}(G) = \{d_1, d_2, ..., d_n\}$, where $d_i = d(v_i)$ is the degree of node $v_i \in V$. Let $\bar{d}$ be the nodes' average number of friends and $\tilde{d}$ the average number of nodes' friends of friends. Then, $\tilde{d} \geq \bar{d}$.

The friendship paradox is, in fact, quite easy to prove. The easiest proof is a "story proof." Suppose all nodes report the how many friends their friends have. How many reports are made? Since each edge in the network connects two nodes, and each edge entails two reports, the total number of reports is $2m$. Next, consider a focal node v with degree d_v. All of v's friends will report that v (their friend) has d_v friends. As there are d_v friends who make the report, there will give a total of d_v^2 reported friends of friends. Summing this over all nodes, we see that $\sum_{v=1}^n d_v^2$ friends of friends are reported. It follows that the average number of reported friends of friends is

$$\tilde{d} = \frac{1}{2m}\sum_{v=1}^n d_v^2.$$

It remains to show that $\tilde{d} \geq \bar{d}$. By the Cauchy-Schwarz Inequality,[1] we have

$$\frac{2m}{n}\tilde{d} = \frac{1}{n}\sum_{i=1}^n d_i^2 \geq \left(\frac{1}{n}\sum_{i=1}^n d_i\right)^2 = \bar{d}^2$$

Dividing both sides of the inequality by $\bar{d} > 0$, and noting that $n\bar{d} = 2m$ (when we sum all degrees, we get two times the total number of edges, since every edge is counted twice), we obtain

$$\tilde{d} \geq \bar{d}$$

[1] The Cauchy-Schwartz Inequality states that for two vectors $v, w \in \mathbb{R}^n$, $|v \cdot w|^2 \leq (v \cdot v)(w \cdot w)$, where $\cdot$ denotes the inner (dot) product. Choosing v to be the degree sequence of G and w to be a vector of ones gives you the inequality above (you'll have to divide both sides by n^2 somewhere...). In fact, the Cauchy-Schwarz Inequality holds not only for $\mathbb{R}$ but all inner product spaces.

which completes the proof. Notice that the Cauchy-Schwarz Inequality will hold as an equality if and only if G is a regular graph, i.e., a graph in which each node has exactly the same degree.[2]

Average Mean Degree of Friends

Feld goes, in fact, one step further to consider the average mean degree of friends—i.e., the mean degree of friends that each individual observes, averaged over all individuals—and argues that this number will be greater than the mean degree.

To prove this claim, consider the graph $G(V, E)$ with associated adjacency matrix $A = (a_{ij})$. Then for each node $v_i \in V$, the total number of her friends' friends is $\tilde{f}_i = \sum_{j=1}^{n} a_{ij} d_j$. As v_i has $d_i = \sum_{j=1}^{n} a_{ij}$ friends, the average number of v_i's friends' of friends is

$$\tilde{f}_i = \frac{t_i}{d_i} = \frac{1}{d_i}\sum_{j=1}^{n} a_{ij} d_j = \frac{1}{d_i}\sum_{j=1}^{n} a_{ij} d_j.$$

Notice that $d_i \geq 1$ for all $i = 1, 2, ..., n$ since the graph is assumed to be connected. Hence, the expression is well-defined. So, the average is

$$\bar{f} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1} a_{ij}\left(\frac{d_j}{d_i}\right) = \frac{1}{n}\sum_{i=1}^{n}\sum_{(i,j)\in E}\left(\frac{d_j}{d_i}\right) \tag{1}$$

since $a_{ij} = 1$ if the edge $(i, j) \in E$. Notice that this sum is over all edges $(i, j) \in E$, where the contribution of each edge (i, j) to the total sum is $d_i/d_j + d_j/d_i$. Compare this expression with the average degree:

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} = \frac{1}{n}\sum_{i=1}^{n}\sum_{(i,j)\in E} 1, \tag{2}$$

which sums over all edges of G as well, but where the contribution of each edge (i, j) is $1 + 1$.

Hence, we see that $\bar{d} \leq \bar{f}$ if

$$2 \leq \frac{d_i}{d_j} + \frac{d_j}{d_i}.$$

And, indeed,

$$\frac{d_i}{d_j} + \frac{d_j}{d_i} = 2 + \frac{(d_i - d_j)^2}{d_i d_j} \geq 2$$

since $d_i > 0$. So we are done.

As in the original friendship paradox, notice that equation (1) will reduce to (2), when $d_i = c$ for all i—i.e., if G is a regular graph.

Generalized Friendship Paradox

The generalized friendship paradox refers to nothing more than the observation that for any node-level characteristic that is positively correlated with degree, a similar result holds: namely, your friends will have more of that characteristic, on average, than you have. To see why this is true, consider a characteristic x. The average x for the nodes and the node's friends is

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \text{and} \qquad \tilde{x} = \frac{\sum_{i=1}^{n} d_i x_i}{\sum_{i=1}^{n} d_i}.$$

[2] In other words, $|v \cdot w|^2 = (v \cdot v)(w \cdot w)$ if and only if $v = \alpha w$, where α is a constant. Thus, $\bar{d} = \tilde{d}$ if and only if the degree sequence and a vector of ones are linearly dependent.

Notice that if $x_i = d_i$, we get back to the original friendship paradox.

We might say that the generalized friendship holds if $\tilde{x} > \bar{x}$ or, equivalently, when $g(\tilde{x}, \bar{x}) = \tilde{x} - \bar{x} > 0$. A bit of algebra reveals that

$$\begin{aligned} g(\tilde{x}, \bar{x}) &= \frac{\sum_{i=1}^{n} d_i x_i}{\sum_{i=1}^{n} d_i} - \bar{x} \\ &= \frac{\sum_{i=1}^{n} (d_i - \bar{d})(x_i - \bar{x}) + n\bar{d}\bar{x}}{n\bar{d}} - \bar{x} \\ &= \frac{\sum_{i=1}^{n} (d_i - \bar{d})(x_i - \bar{x})}{n\bar{d}} \\ &= \frac{\mathrm{Cov}(d, x)}{\bar{d}}. \end{aligned}$$

As $\bar{d} > 0$, $g(\tilde{x}, \bar{x}) > 0$ if $\mathrm{Cov}(x, d) > 0$, which shows that the generalized friendship paradox will hold for any characteristic that is positively correlated with degree.