A Surprising Fact About Friendship

http://csustan.csustan.edu/~ tom

Tom Carter

Computer Science
CSU Stanislaus

tom@csustan.csustan.edu

http://csustan.csustan.edu/~
tom/Lecture-Notes/Friendship-paradox/Friendship-paradox.pdf

September 29, 2014

Our general topics:

Expected values and averages $\leftarrow$

Suppose we have a *probability distribution* $\mathcal{P} = \{p(i) | i \in I\}$, where I is a finite *index set*, with $0 \leq p(i) \leq 1$, and $\sum_{i \in I} p(i) = 1$. We think of the probabilities $p(i)$ as telling us the probability that index i in the *index set* I will occur.

Now, suppose that we have some function $X : I \to \mathbb{R}$, which gives a particular value $X(i)$ whenever index i occurs. We call X a *random variable* over the probability distribution $\mathcal{P}$. We can then calculate the *expected value* (or the *mean*) of the random variable X as:

$$< X >= \sum_{i \in I} X(i) p(i).$$

The *expected value* has some nice properties.

Suppose X and Y are random variables over the same probability distribution $\mathcal{P}$, and $c \in \mathbb{R}$. Then we have:

$$< X + Y >=< X > + < Y >$$

and

$$< cX >= c < X >$$

(in other words, *expected value* is *linear*). These properties follow easily from the definition.

We also have the obvious property that if $X(i) = c$ for all i, then

$$< X >=< c >= c.$$

This also implies that, since $< X >$ is independent of i, we have

$$<< X >>=< X > .$$

The *first moment* (often denoted μ) is the expected value of X:

$$\mu =< X > .$$

The *second moment* (usually called the *variance*, or σ^2) is:

$$\sigma^2(X) = Var(X) =< (X - \mu)^2 > .$$

We can do the quick calculation:

$$\begin{aligned} Var(X) &= \sigma^2(X) \\ &=< (X - \mu)^2 > \\ &=< X^2 - 2X\mu + \mu^2 > \\ &=< X^2 > - < 2X\mu > + < \mu^2 > \\ &=< X^2 > -2\mu < X > +\mu^2 \\ &=< X^2 > -2 < X >^2 + < X >^2 \\ &=< X^2 > - < X >^2 . \end{aligned}$$

Average number of friends $\leftarrow$

Let's do some calculations about an "acquaintanceship" network . . .

Let $G = (V, L)$ be an undirected graph (an "acquaintanceship" network) where V is a (finite) set of vertices (or nodes), and L is a set of unordered pairs of vertices from V (links, or edges). We will assume there are no self-edges (i.e., if $(v_i, v_j) \in L$, then $v_i \neq v_j$), and no duplicate edges. For simplicity, we can assume $i < j$. Let $N = |V|$, the number of vertices, and $E = |L|$, the number of edges.

For this example, we will be interested in the *degrees* of the vertices. In particular, for each $v \in V$, the degree of a vertex v is

$$d_1(v) = |\{(v_i, v_j) | v = v_i \text{ or } v = v_j\}|,$$

This is the number of vertices that can be reached in 1 step from v, or the number of paths of length one leaving v.

Note that we will have

$$\sum_{v \in V} d_1(v) = 2E$$

since each edge contributes to the degrees of two vertices.

Let's calculate the average degree of vertices in the graph. Let D_1 be the random variable corresponding with the degree of a vertex v chosen (uniformly) randomly (that is, the number of paths of length one leaving v).

The average degree will be:

$$\begin{aligned} < D_1 > &= \sum_{v \in V} d_1(v) p(v) \\ &= \sum_{v \in V} d_1(v) \frac{1}{N} \\ &= \frac{1}{N} \sum_{v \in V} d_1(v) \\ &= \frac{1}{N} * 2E \\ &= \frac{2E}{N} \end{aligned}$$

Note that $p(v) = \frac{1}{N}$ is the (uniform) probability of choosing vertex v.

(This result is no surprise at all . . . :-)

Friends of friends $\leftarrow$

Let's generalize this a bit. Let's look at the distribution of degrees of vertices that are one step away from a given vertex. In other words, we're looking at the distribution of how many friends your friends have. We'll do this by looking at paths of length two.

Let D_2 be the random variable expressing the number of vertices connected to a vertex by paths of length two – that is, the number of paths of length two leaving a vertex.

Let $p_1(i)$ be the distribution of degrees:

$$p_1(i) = \frac{n_i}{N}$$

where n_i is the number of vertices of degree i, and N is the total number of vertices.

Now we calculate $p_2(i)$, the probability that starting at a random vertex, there are i paths of length 2 leaving that vertex. First we calculate the total number N_2 of paths of length 2. We can do this by summing over the possible degrees of the vertices in the middle of such paths, and observing that each node of degree i is in the middle of i^2 paths of length 2:

$$\begin{aligned} N_2 &= \sum_i i^2 n_i \\ &= \sum_i i^2 N p_1(i) \\ &= N \sum_i i^2 p_1(i) \\ &= N < {D_1}^2 > \end{aligned}$$

This means that the average number of paths of length 2 starting at a uniformly randomly chosen vertex will be

$$
\begin{aligned}
< D_2 > = \frac{N_2}{N} = \frac{N < {D_1}^2 >}{N} \\
= < {D_1}^2 > .
\end{aligned}
$$

From this we can calculate the average number of friends a friend has (on average):

$$
\begin{aligned}
\frac{\text{average number of friends of friends}}{\text{average number of friends}} = \frac{< D_2 >}{< D_1 >} \\
= \frac{< {D_1}^2 >}{< D_1 >}
\end{aligned}
$$

We can do some minor calculations with this:

$$\begin{aligned}\frac{< {D_1}^2 >}{< D_1 >} &= < D_1 > + \frac{< {D_1}^2 >}{< D_1 >} - < D_1 > \\ &= < D_1 > + \frac{< {D_1}^2 > - < D_1 >^2}{< D_1 >} \\ &= < D_1 > + \frac{Var(D_1)}{< D_1 >}\end{aligned}$$

Note that since $Var(D_1) = \sigma^2(D_1) \geq 0$, this means that, unless everybody has exactly the same number of friends, then

On average, your friends have more friends than you do . . .

One further observation . . .

We have that the excess average degrees of neighbors (on average) is

$$\frac{Var(D_1)}{< D_1 >}$$

Thus, the larger the variance, the larger the excess. So, for example, if the degree distribution is (approximately) a *power law*, we can expect the variance to be large . . .

Notes $\leftarrow$

I'm leaving a few notes here, potentially for future reference . . .

We have that $p(i)$ is the distribution of degrees, or paths of length 1. As noted, we have: $p(i) = \frac{n_i}{N}$.

We also want to look at the distribution of paths of length 2. Let $p_2(i)$ be the probability that, starting at a random vertex, there are i paths of length 2 leaving that vertex. We may be able to calculate this by first looking at the probability $q(i)$ of arriving at a degree i node by following an edge. We can calculate this by calculating how many ends of edges are at vertices with degree i, and dividing by the total number of ends of edges. We will have:

$$\begin{aligned} q(i) &= \frac{\text{number of ends at degree } i \text{ vertices}}{\text{total number of ends of edges}} \\ &= \frac{i * (\text{number of degree } i \text{ vertices})}{\text{total number of ends of edges}} \\ &= \frac{i * (N * p(i))}{2E} \\ &= i * \frac{N}{2E} * p(i) \\ &= i * \frac{1}{< D_1 >} * p(i) \\ &= \frac{i}{< D_1 >} p(i). \end{aligned}$$

Quick check: note that $\sum q(i) = 1$, as needed.

Another possible approach: let $C_1(v)$ be the set of neighbors of v, and $C_2(v)$ be the set of neighbors of neighbors of v. Then

$$\begin{aligned} d_2(v) &= |C_2(v)| \\ &= \sum_{w \in C_1(v)} d_1(w) \end{aligned}$$

We then have

$$\begin{aligned} < D_2 > &= \sum_i |C_2(v_i)| p(i) \\ &= \sum_i \sum_{w \in C_1(v_i)} d_1(w) p(i) \\ &= \sum_i d_1(v_i)^2 p(i) \\ &= < {D_1}^2 > \end{aligned}$$

And some more rough notes . . .

We have $p_1(i)$ = distribution of degrees.

Now look at edges:

Let $q_1(i)$ be the probability distribution of arriving at a degree i node by following an edge. We will have:

$$q_1(i) = \frac{i}{\sum_j j p_1(j)} p_1(i) = \frac{i p_1(i)}{< D_1 >}$$

note $\sum_i q_1(i) = \sum_i \frac{i p_1(i)}{<D_1>} = 1$ (as required)

Now we can look at the average number of neighbors of neighbors, per neighbor:

$$\begin{aligned}
\frac{< D_2 >}{< D_1 >} &= \sum_i i q_1(i) \\
&= \sum_i i \frac{i p_1(i)}{< D_1 >} \\
&= \sum_i \frac{i^2 p_1(i)}{< D_1 >} \\
&= \frac{< {D_1}^2 >}{< D_1 >}
\end{aligned}$$

←

References

[1] 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

[2] Newman, M.E.J. (2002),
Random Graphs as Models of Networks,
SFI WORKING PAPER: 2002-02-005,
http://www.santafe.edu/media/workingpapers/02-02-005.pdf

To top ←