Friends of Friends

Yash Gupta — IIT Bombay

Soumen Chakrabarti — IIT Bombay

A well-known 'paradox' in (social) networks is informally stated as "your friends are more popular than you are, on average". A first step to make this eye-catching statement more rigorous is to restate it as "the average degree of a randomly selected node in a network is less than the average degree of neighbors of a randomly selected node", but this leaves unspecified the exact mechanism of averaging. There are at least two interpretations, one well-known, the other, apparently not.

Given an undirected graph $G = (V, E)$ with $|V| = N$ and $|E| = M$ and no isolated nodes.

The set of neighbors of node u is $\mathrm{nbr}(u)$.

The average degree is $\mu = \frac{1}{N} \sum_{u \in V} |\mathrm{nbr}(u)| = \frac{2M}{N} \geq 1$.

Let the number of "friends of friends" of node u be denoted $\mathrm{FF}(u) = \sum_{v \in \mathrm{nbr}(u)} |\mathrm{nbr}(v)|$.

We have $\mathrm{FF}(u) \geq |\mathrm{nbr}(u)| \geq 1$.

Consider the quantities

$$\mathrm{MicroAvg} = \frac{\sum_{u \in V} \mathrm{FF}(u)}{\sum_{u \in V} |\mathrm{nbr}(u)|} \qquad \text{and} \qquad \mathrm{MacroAvg} = \frac{1}{N} \sum_{u \in V} \frac{\mathrm{FF}(u)}{|\mathrm{nbr}(u)|}$$

Claim 1. $\mathrm{MicroAvg} \geq \mu$.

Consider the numerator $\sum_{u \in V} \mathrm{FF}(u) = \sum_{u \in V} \sum_{v \in \mathrm{nbr}(u)} |\mathrm{nbr}(v)|$. This is effectively an enumeration over edges, accumulating on both incident nodes:

```
for each node u ∈ V do
    initialize FF(u) ← 0
for each edge {u, v} ∈ E do
    FF(u) ← FF(u) + |nbr(v)|
    FF(v) ← FF(v) + |nbr(u)|
```

Note that v contributes the term $|\mathrm{nbr}(v)|$ to each of its $|\mathrm{nbr}(v)|$ neighbors, which then add up in $\sum_{u \in V} \mathrm{FF}(u)$. Thus the total contribution of v to the numerator is $|\mathrm{nbr}(v)|^2$. Thus we get

$$\mathrm{MicroAvg} = \frac{\sum_{u \in V} |\mathrm{nbr}(u)|^2}{\sum_{u \in V} |\mathrm{nbr}(u)|} = \frac{\sum_{u \in V} |\mathrm{nbr}(u)|^2}{2M} = \frac{\sum_{u \in V} |\mathrm{nbr}(u)|^2}{N} \frac{N}{2M}.$$

The standard deviation of node degree is defined as

$$\sigma^2 = \frac{\sum_{u \in V} |\mathrm{nbr}(u)|^2}{N} - \mu^2 \quad \Longrightarrow \quad \frac{\sum_{u \in V} |\mathrm{nbr}(u)|^2}{N} = \sigma^2 + \mu^2$$

from which we get

$$\text{MicroAvg} = \frac{\sigma^2 + \mu^2}{\mu} = \frac{\sigma^2}{\mu} + \mu \geq \mu.$$

Claim 2. MacroAvg $\geq \mu$.

Here the computation has to be organized slightly differently.

for each node $u \in V$ **do**
  initialize $Q(u) \leftarrow 0$
for each edge $\{u, v\} \in E$ **do**
  $Q(u) \leftarrow Q(u) + \dfrac{|\operatorname{nbr}(v)|}{|\operatorname{nbr}(u)|}$
  $Q(v) \leftarrow Q(v) + \dfrac{|\operatorname{nbr}(u)|}{|\operatorname{nbr}(v)|}$
return $\frac{1}{N} \sum_{u \in V} Q(u)$

So this time, each edge contributes to MacroAvg the quantity $\dfrac{|\operatorname{nbr}(v)|}{|\operatorname{nbr}(u)|} + \dfrac{|\operatorname{nbr}(u)|}{|\operatorname{nbr}(v)|} \geq 2$, because $\min_{a,b>0} \frac{a}{b} + \frac{b}{a} = 2$. We thus get

$$\text{MacroAvg} = \frac{1}{N} \sum_{u \in V} Q(u) \geq \frac{1}{N} \cdot M \cdot 2 = \frac{2M}{N} = \mu.$$

The remaining question is whether some inequality holds between MacroAvg and MicroAvg.

Claim 3. MicroAvg > MacroAvg is possible.

Consider a 4-node clique on nodes $1, 2, 3, 4$, followed by a chain $4, 5, 6$.

Node u	$\operatorname{nbr}(u)$	$\operatorname{FF}(u)$	$\operatorname{FF}(u)/|\operatorname{nbr}(u)|$
1	2,3,4	3+3+4=10	$10/3 = 3.\dot{3}$
2	1,3,4	3+3+4=10	$10/3 = 3.\dot{3}$
3	1,2,4	3+3+4=10	$10/3 = 3.\dot{3}$
4	1,2,3,5	3+3+3+2=11	$11/4 = 2.75$
5	4,6	4+1=5	$5/2 = 2.5$
6	5	2	$2/1 = 2$
	$\mu = 16/6 = 2.\dot{6}$	$\sum_u \operatorname{FF}(u) = 48$	$\sum_u \frac{\operatorname{FF}(u)}{|\operatorname{nbr}(u)|} = 17.25$

$\text{MacroAvg} = \frac{1}{N} \sum_u \frac{\operatorname{FF}(u)}{|\operatorname{nbr}(u)|} = \frac{17.25}{6} = 2.875 < \text{MicroAvg} = \frac{48}{16} = 3$

Claim 4. MicroAvg < MacroAvg is possible.

Consider the square on nodes $1, 2, 3, 4$ with diagonal $2, 4$ connected.

Node u	$\operatorname{nbr}(u)$	$\operatorname{FF}(u)$	$\operatorname{FF}(u)/|\operatorname{nbr}(u)|$
1	2,4	3+3=6	6/2=3
2	1,3,4	2+2+3=7	$7/3 = 2.\dot{3}$
3	2,4	3+3=6	6/2=3
4	1,2,3	2+3+2=7	$7/3 = 2.\dot{3}$
	$\mu = 10/4 = 2.5$	$\sum_u \operatorname{FF}(u) = 26$	$\sum_u \frac{\operatorname{FF}(u)}{|\operatorname{nbr}(u)|} = 10.\dot{6}$

$\text{MacroAvg} = \frac{1}{N} \sum_u \frac{\operatorname{FF}(u)}{|\operatorname{nbr}(u)|} = \frac{10.\dot{6}}{4} = 2.\dot{6} > \text{MicroAvg} = \frac{26}{10} = 2.6$