Do your friends have more friends than you do?

Davide Baldelli

Alma Mater Studiorum - University of Bologna, Bologna, Italy
davide.baldelli4@studio.unibo.it
ID: 0001052636

Abstract. The friendship paradox is the phenomenon discovered by Scott L. Feld [1], according which most people have fewer friends than their friends have, on average. In spite of its apparently paradoxical nature, it can be proved mathematically that in a social network (modeled as an undirected graph), the average number of friends of individuals is lower than the average of the mean number of the persons' friends' friends. Anyway, it does not prove that in average your friends have more friends than you do, as one should prove that the proportion of individuals who have fewer friends than the mean number of friends their own friends have is higher than fifty percent. I am going to search and explore networks in which most of the nodes will have more friends than the mean number of its friends' friends, to understand if that mathematical result is sufficient to conclude that most of the individuals will have fewer friends than their friends have.

1 Introduction

The main idea behind the paper published by Scott L.Feld, Why Your Friends Have More Friends than You Do [1], is that if individuals use the number of friends that their friends have as one basis for determining whether they have an adequate number of friends it is likely that most of them will feel inadequate. It has been proven mathematically that, assuming that a social network is represented by an undirected graph $G = (V, E)$, where the set V of vertices corresponds to the people in the social network, and the set E of edges corresponds to the friendship relation between pair of nodes, the average number of friends of a person is lower than the average of the mean number of the person's friends' friends.
To say it formally, if we define respectively x_i and y_i as the number of friends and the average number of friends' friends of the i^{th} node, N as the number of nodes, and I as the set of individuals, the average number of friends of each individual is simply:

$$AVG_f = \mu(x) = \frac{\sum_{i \in I} x_i}{N} \tag{1}$$

where with μ I mean the expected value of a probabilistic distribution. While the average number of friends' friends of an individual is

$$AVG_{ff} = \mu(y) = \frac{\sum_{i \in I} x_i^2}{\sum_{i \in I} x_i} \tag{2}$$

This equation can be shown to be equal to

$$AVG_{ff} = \mu(x) + \frac{\sigma^2(x)}{\mu(x)} \tag{3}$$

that proves that $AVG_{ff} \geq AVG_f$.
With this paper I want to argue that this mathematical result does not prove that most of individuals have fewer friends that their friends have, because the mathematical projection of this phenomenon is that on average (between different networks) the following condition is satisfied:

$$|\{v_i \in V \quad s.t. \quad x_i > y_i\}| > \frac{N}{2}$$

Despite this fact, it is worth noting that the mathematical result makes it reasonable to suppose that in many scenarios most of individuals will have fewer friends than their friends have. The intuition behind it is that if in a network most of the nodes have $x_i > y_i$ but $AVG_{ff} > AVG_f$, the network must have a distribution of x and y not *reflecting* the trend of the averages, and thus having a particular topology. As a consequence, I am going to simulate several situations to explore when and why networks in which most of the nodes have more friends than their friends do form.

2 Experiments

I have conducted the experiments using NetLogo [3]. Four different experiments have been developed. In all the experiments the networks have been initialized with the Erdos-Renyi model [2], since I wanted to conduct the exploration with the simplest and the most random network's configurations.

2.1 Metrics

To simplify the notation, I define as **happy node** a node that has more friends than the average number of its own friends' friends. For the sake of disambiguation, we define as happy nodes only the ones whose number of friends is strictly greater than the average number of its friends' friends, and when it happens that a node does not have any friend, we do not record its y_i. In other words, AVG_{ff} is computed without taking into consideration singlets.
The following metrics have been recorded during the experiments:

- Ratio between AVG_{ff} and AVG_f;
- Ratio between the amount of happy nodes and the total number of nodes:

$$\frac{|\{v_i \in V \quad s.t. \quad x_i > y_i\}|}{N}.$$

2.2 Experiments' design

Experiment 1. In the first experiment I am going to explore how the ratio of happy nodes, and the ratio between AVG_{ff} and AVG_f varies taking different wiring probabilities p when creating random networks with the Erdos-Renyi model. I have run the experiment for different number of nodes: $10, 25, 50, 75, 100$, and 200. For each number of nodes I have created 100 different networks for each p varying between 0 and 1, with an increasing step of 0.01. For each case, I have recorded the metrics and I have taken the average between the 100 different observations, in order to reduce the noise consequent to the peculiar randomness of the Erdos-Renyi model.

Experiment 2. Here we start exploring how the situation changes adding some dynamics. In this experiment a random network with a fixed number of nodes is initialized and a random edge is added at each step until the network becomes completely connected. The experiment has been conducted with $10, 25$ and 50 nodes. Here I have exploited the variant of the Erdos-Renyi model in which we fix the initial number of edges, and I have set this parameter to be the expected value μ_E of number of edges if we initialize the network with the classic Erdos-Renyi model with $p = 0.1$:

$$\mu_E = \binom{N}{2} \cdot p \tag{4}$$

For each number of nodes, I have run 10 experiments.

Experiment 3. In this experiment a random network with the Erdos-Renyi model is created and edges are added and deleted in a noisy way: at each time step there is 50% of probability that edges will be added or deleted and the amount of those edges is extracted from a uniform random distribution.
Furthermore, in the process of adding new edges, a node that has more friends than a certain threshold m, could not create new relationships. I have run the experiment in two different conditions, with the main constraint that the initial average number of friends per node, μ_e, should be higher than the maximum allowed relationships m, in order to see the evolution of the metrics while AVG_f converges to that parameter. The average number of friends per node could be deduced by equation 4 and is expressed by the following equation:

$$\mu_e = \mu_E \frac{2}{N} = (N-1) \cdot p \tag{5}$$

In both the configurations, the number of nodes is set to 50 and the maximum number of added or deleted edges at each time step is 4. After that:

- In the first configuration, $p = 0.3$, $\mu_e = 14.7$, and $m = 5$.
- In the second one, $p = 0.4$, $\mu_e = 19.6$, and $m = 10$;

Experiment 4. Here I have implemented two *ad hoc* dynamical processes with the aim of increasing the average happiness of the network. In particular, at each time step the same noisy edge generation and deletion process of the previous experiment is computed, with the maximum number of possibly added or deleted edges set to 2. In addition the following two procedures are run:

- an edge between two random unhappy nodes is created;
- an edge between the most popular node (the one with the highest number of friends) and its most popular friend is deleted.

The number of nodes is set to 100 and the initial network is created with the Erdos-Renyi model with $p = 0.1$. The experiment has been run for 10 times, each lasting 100 time steps.

3 Results

Experiment 1. It is reasonable to suppose that a randomly initialized network would not have an enough sophisticated topology to diverge from what predicted by the inequality between AVG_{ff} and AVG_f. I would also say that a sparser random network will have an higher variance in the distribution of edges and, as a consequence, an higher ratio between AVG_{ff} and AVG_f and a lower ratio of happy nodes. While, increasing p both AVG_f and AVG_{ff} will converge to $N-1$, their ratio will approach 1 and the ratio of happy friends will increase.
As we can see in Figure 1 the simulations fulfill our predictions. Furthermore, we can notice that higher the number of nodes, faster will be the convergence between AVG_{ff} and AVG_f and the convergence of the ratio of happy nodes to 0.5. We can observe from the third graph in figure that for very highly connected networks ($p > 0.9$) the ratio of happy nodes is mostly higher than 50%, and, while p approaches 1, the ratio follows an almost chaotic behaviour. The reason behind this could be that in very connected networks x_i and y_i are almost $N-1$ for each node, and slight differences in the topology of the configuration can easily turn many nodes to be happy or unhappy.

Experiment 2. Similarly to *Experiment 1* in this experiment we go from sparse to highly connected networks. Indeed, the results are similar: the ratio between AVG_{ff} and AVG_f converges to 1 while the ratio of happy nodes approaches 50% (see Figure 2). While converging to 50% the ratio of happy nodes oscillates with an increasingly wider amplitude. This phenomenon could be explained in a similar way than for *Experiment 1*, i.e. it could be that as AVG_f and AVG_{ff} converge, most of the nodes would have $x_i = y_i$, and the creation of a single edge can make many nodes happy or unhappy. The reason why the ratio of happy nodes oscillates in a orderly way is not completely understood.

Experiment 3. We expect the networks' nodes in this experiment to become happier as AVG_f approaches m, for the same reason of the first two experiments: a lower variance of the distribution of x makes more nodes happy.

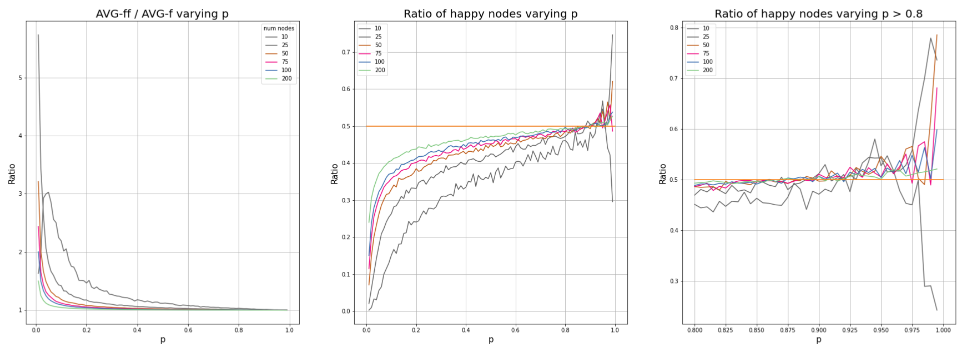
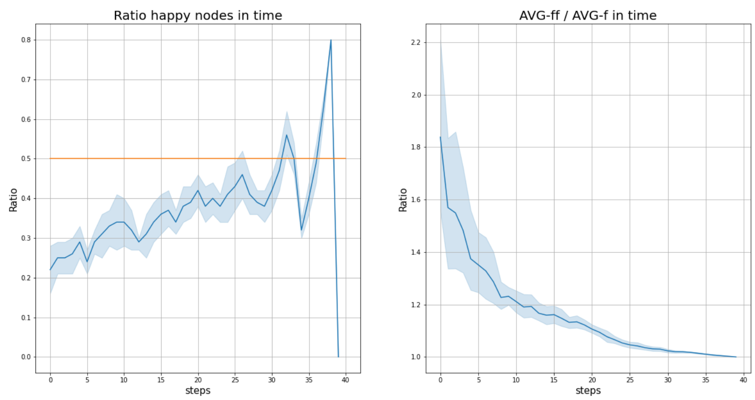


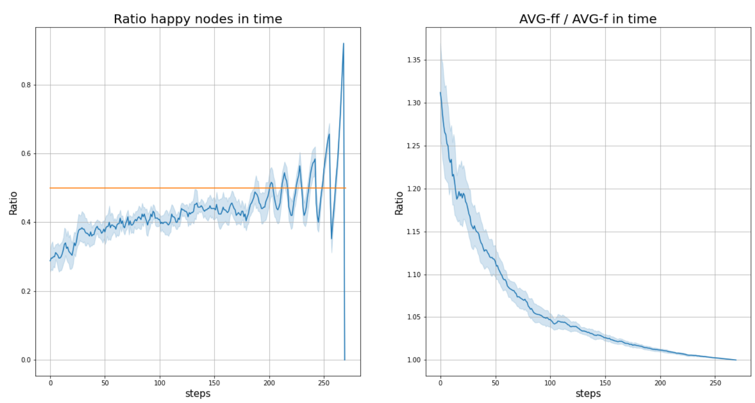
Fig. 1: Results of Experiment 1. In the first graph is shown the trend of the ratio AVG_{ff}/AVG_f varying p. In the second and the third one is shown the ratio of happy nodes varying p. In the third graphic is shown the trend varying p in the range $[0.8, 1]$.

Indeed, we can observe in Figure 3 that in in both cases, as soon as AVG_f reaches the value of m (approximately at step 300 in both cases), the ratio of happy nodes start increasing sharply. In the second case it seems that it stabilizes in a generally happy network, while in the first case the ratio does not exceed 0.5 definitively. Why it has happened is not completely clear, but it leaves the possibility for happier networks to be discovered and studied.
On the other hand, the ratio between AVG_{ff} and AVG_f starts increasing when AVG_f reaches m in the first case and it stabilizes in the second one. I would have expected that in both cases the ratio would have started decreasing, thus the observed behaviour is not easily explainable.

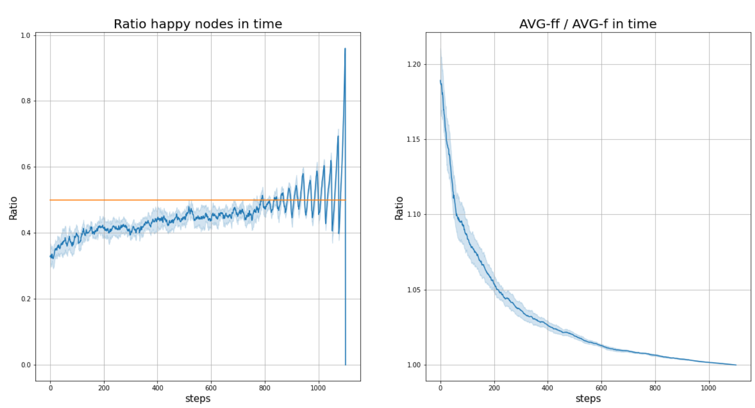
Experiment 4. In the last experiment the procedures to increase the average happiness reveal to be effective despite of the noisy edge creation and deletion: as we can see in Figure 4, in just 100 steps the network passes from a ratio of happy nodes lower than 0.4, to a ratio higher than 0.55. As noted in all the experiments, an increase in the ratio of happy nodes corresponds to a convergence of AVG_{ff} and AVG_f. In this case AVG_f stays stable while AVG_{ff} approaches AVG_f. It can be explained by noticing that the two procedures tend to keep the number of friends of each node equal by deleting a node between very happy nodes and creating a node between unhappy nodes, and, doing so, they flatten inequalities, making AVG_{ff} evolve towards AVG_f.



(a) 10 nodes.

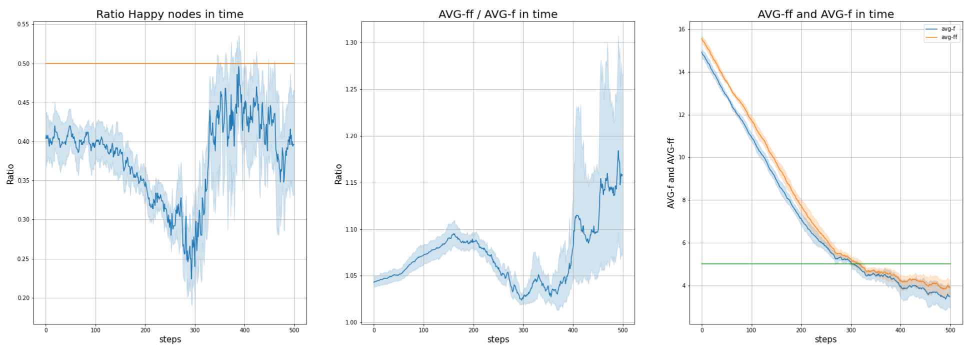


(b) 25 nodes.

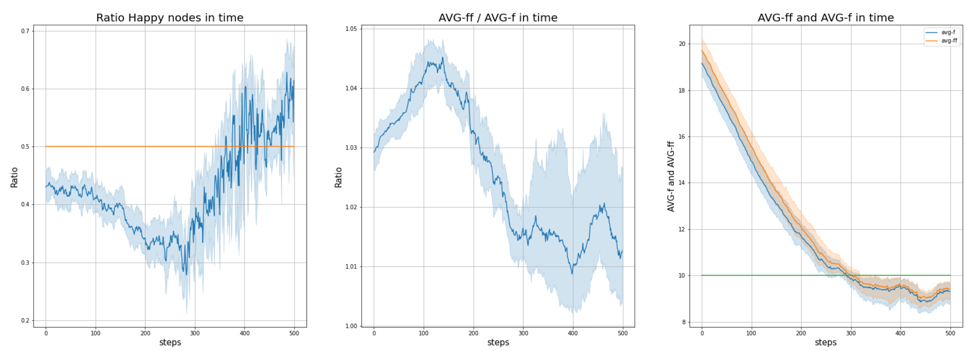


(c) 50 nodes.

Fig. 2: Results of Experiment 2. The pairs of graphs show respectively the trend of the ratio of happy nodes and the trend of the ratio between AVG_{ff} and AVG_f for different number of nodes. The lighter zones are the 95% confidence intervals computed between the 10 experiments.



(a) $p = 0.3, m = 5$.



(b) $p = 0.4, m = 10$.

Fig. 3: Results of Experiment 3. In (a) the results of the first case, in (b) the results of the second case. The green horizontal line in the graphs of the trend of AVG_f and AVG_{ff} indicates m. The lighter zones are the 95% confidence interval computed between the 10 experiments.

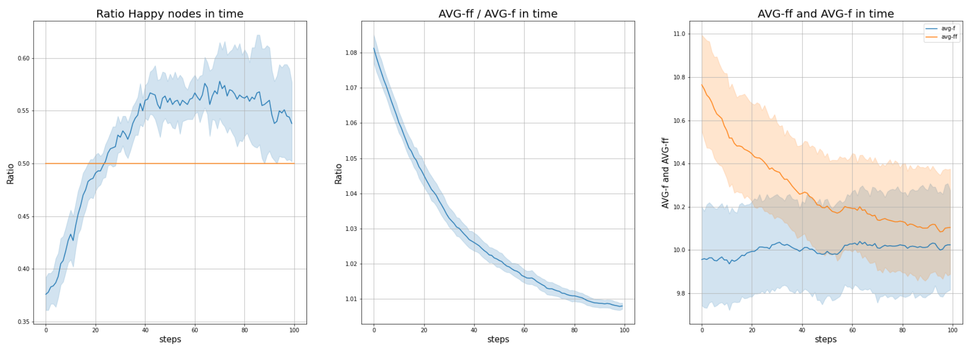


Fig. 4: Results of Experiment 4. It is shown respectively: the ratio of happy nodes in time, The ratio between AVG_{ff} and AVG_f in time, and the trends of AVG_{ff} and AVG_f. The lighter zones are the 95% confidence intervals computed between the 10 experiments.

4 Conclusions

From all the experiments emerges that typically an increase in the number of happy nodes corresponds to a convergence between AVG_f and AVG_{ff}. At this point is not clear if it is a necessary requirement for the ratio of happy nodes to be high.
As expected, in random configurations the ratio of happy nodes exceeds 50% only in the most *pathological* conditions, that is to say in very highly connected networks.
In experiments 3 and 4 I have shown that this highly connection is not a necessary condition for happy network to emerge. Indeed, in experiment 3 the number of edges decreases during the dynamic and the ratio of happy nodes increases sharply, while in experiment 4 the number of edges stays stable around the initial value.
In conclusion, do your friends have more friends than you do? My study provides evidence that it is possible to build network in which this statement is not true anymore, despite the fact that $AVG_{ff} > AVG_f$. However, my study also highlights the need for further research in this area. For example, the study only examined artificially built networks, so it is unclear whether our findings can be generalized to actual data.

References

1. Feld, Scott L. (1991), "Why your friends have more friends than you do", American Journal of Sociology.
2. Erdos, P.L., Renyi, A. (1984). On the evolution of random graphs. Transactions of the American Mathematical Society, 286, 257-257.

3. Wilensky, U. (1999). NetLogo. http://ccl.northwestern.edu/netlogo/. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.