Friendship network analysis

Isabella Marasco (1040993)
isabella.marasco3@studio.unibo.it

University of Bologna

Abstract. In this document is proposed a model to better understand the phenomenon of friendships. It happens very often that individuals use their number of friends to understand how adequate they are in their social circle. The aim is to understand how as the reputation of individuals changes, this can also affect the relationships and popularity of an individual.

1 Introduction

Society has always been composed of individuals with networks of friendships. Friendship is not only a source of satisfaction, support, and sharing but often, especially in modern society, it becomes a way of comparing oneself to others. Many people compare their own number of friends with their friends' friends to measure how adequate they are in their social circle, and most likely most people who make this kind of comparison feel inadequate [1]. This happens because a person with few friends is more likely to have a friend with more friends than him. In addition, a person's reputation can also affect his or her friend's reputation, which is why people very often tend to exclude those who do not have a good reputation from their social circle.

1.1 Behavior in friendships

Various theories have been proposed to try to explain the behavior and reasons why an individual will always have fewer friends than their friends.
One of the theories used is the one found within the paper, from which this project was inspired, Why Your Friends Have More Friends than You Do [1] in which it is stated that the average number of friends of friends is always greater than the average number of an individual's friends, this happens because it is easier to know people with more friends. Further analysis shows that the proportion of individuals who have fewer friends than the mean number of friends their own friends have is affected by the exact arrangement of friendships in a social network.
In addition to these theories from which to carry out this project, another study to be carried out will be to see how the number of friends varies as popularity increases and how the popularity is also affected as an individual's reputation changes.

2 Model definition

The network consists of several agents whose relationships are represented by the links to other agents on the network, and its structure was realized by implementing the Erdős-Rényi model.
The proposed model was implemented using two types of approaches. The first based on the infection diffusion, in which agents may be infected, i.e. because of a bad reputation that has spread about an agent. The infection could spread to healthy agents linked with the infected ones, so to avoid it, one could decide to eliminate the link with the infected agent or simply wait for the infection to pass. The second approach is game theory, in which each individual decides which strategy to use, that is, whether to become friends with another individual based on his or her popularity. These two approaches can also be used together.

2.1 Environment population

The number of agents present within the environment is defined by the user by the parameter *population*.
The initial composition of the population, to define the number of *healthy* and *infected* agents, can be changed using the *initial-infected* parameter, which to define the number of initially infected agents is defined.
Initially, the population is arranged in a circular fashion on the map and the infected are randomly placed among the healthy ones.

2.2 Network model

Network modeling was implemented using the Erdős-Rényi model, which can be obtained from two different possible choices for creating the arcs:

- Randomize the probability of an arc being formed between two nodes by entering it from input using the parameter p.
- Enter from input the number of links within the network using the *num-links* parameter.

2.3 Interactions

Interactions between individuals can change in the time and there can be a multitude of causes. New friendships can be established or eliminated due to a bad reputation or for other reasons.
To simulate such occurrences within the project, some possible interactions that could take place between agents were implemented:

- Randomly modify their interactions, via the *update-network-random* button with which, one enters from input the probability of a friendship being deleted using the *drop-friendship-p* parameter, otherwise a new connection is added.

- Change their interactions via the *add or delete infected links* button, by which, you enter from input the probability that a new friendship will be added using the *add-friendship-p* parameter. In case you are within that probability you create new friendship only with uninfected people otherwise drop friendship with infected people. This interaction is possible only if *infections* switch is *On*.

2.4 Infection diffusion model

Another interaction between individuals can be simulated through the *infection diffusion model*, which is possible if there is at least one individual within the network who, for whatever reason, has a bad reputation that could affect his or her connections as well. The execution of this interaction is possible if the *infection* switch is set to *On*.
To realize this interaction initially all individuals are set as *healthy* while from input with the *initial-infected* parameter the number of infected agents are defined. After this is done, it are set the parameter from input to indicate the probability with which the virus is able to spread more or less quickly, that individuals will be exposed to the virus, that an infected individual will become healthy again, and finally the frequency with which to check for the presence or absence of the virus can be established, respectively with *virus-spread-chance*, *exposed-chace*, *recovery-chance*, and *virus-check-frequency*.
Each agent can be in four states:

- ***Healthy***: is the state in which all agents are initially set, meaning that they have never been infected.
- ***Infected***: indicates the agents that are infected at that time in the simulation and therefore can infect others connected to it.
- ***Recovered***: indicates the individual who after being infected has been cured, for example, a individual who has managed to get rid of a bad reputation.
- ***Exposed***: the individual who have connections with infected people are considered as such, meaning that although they are not yet infected they could become infected.

2.5 Game theory

The last possible interaction implemented within the project is based on the game theory model, the execution of which is possible only if the *game-theory* switch is set to *On*.
Each agent is considered popular or unpopular in relation to the value of the popular parameter entered from input. Each individual can interact with others in the network by going to compare their popularity, if the two agents are both popular then they will become friends otherwise no friendship will be added. The payoff matrix can be observed below.

	Popular	**No popular**
Popular	1	0
No popular	0	1

Table 1. Payoff matrix

2.6 Combination Diffusion Infection and Game theory

Setting both the Infection and *game-thoery* switches to *On* allows the two interaction models to be executed simultaneously.
In this case the execution of *infection-thoery* does not change while, as far as game theory is concerned the basic execution remains unchanged, i.e., two individuals become friends only if both are popular, with the activation of *infection-thoery* we check that both agents are popular but will form a link only with agents that are not infected.

3 Model experiments

The main interest is to understand how as interactions between different types of agents (healthy, exposed, infected) change the number of friends of the individual. Below, we will look at the simulations obtained from the model and the results obtained: The metrics used in the project are:

- ***AVG_f*** : the average number of friends of each individual.

$$AVG_f = \frac{\sum_{iI} x_i}{N}$$

with xi being the number of friends of the individual i, N the total number of individuals and I the set of individuals.
- ***AVG_ff*** : the average number of friends' friends of an individual.

$$AVG_{ff} = \frac{\sum_{iI} x_i^2}{\sum_{iI} x_i}$$

since an individual has xi friends and contributes to each of its friends average as well.
- ***Cluestering coefficient***: measure of how closely its neighbors are connected. It can take on any value between 0 and 1.0. A value of 0 indicates that none of the neighbors of the node are connected to each other, while a value of 1.0 indicates that all of the neighbors are connected to each other.

In the following section we'll see how well the model performs and what kind of insights we can get from its execution. Each experiment will be run for 1000 steps with a population of 200 individuals.

3.1 First experiment

Infection diffusion

The network was created using the parameter p set to 0.30. The number of infected is 120 in the population, the probability of *virus-spread-chace* is 0.7, *exposed-chance* is 0.8 while *recovery-chace* is 0.4.
The simulation was realized by running the infection diffusion, integrating other interactions using *update-network-random*, and setting the probability of dropping friendships, using the *drop-friendship-p* parameter, equal to 0.2.
During the simulation it can be seen that the number of infected initially is quite high, this with the passage of time tends to decrease, a similar trend can be seen in the exposed that after an increase in a fast way starts to decrease, when it comes to recovered one can observe a continuously increasing trend while, the number of healthy has a net decrease since the beginning of the simulation, surely due to the high number of initial infected and therefore a high number of exposed from the beginning.
These trends are caused by several factors, the network is not tightly connected so one would think that the infection could spread with more difficulty but by containing many infected individuals from the start and having high spread and exposed probabilities it makes sense to get a network with many exposed especially in the beginning.
The average number of friends (AVG_f) and the average number of friends of friends (AVG_ff) both increased, although slightly, because by setting the *drop-friendship-p* parameter to a low value the network tends more to increase the number of friendships and not to drop them. Despite these changes, the value of AVG_f always remained lower than AVG_ff.
The clustering coefficient initially has a value of 0.3 equal to the probability that a friendship is created between two individuals. This happens because the network links were made using the p parameter by creating a link between two random agents with that probability, regardless of their context. The clustering coefficient increased slightly due to the interaction of the agents and the increase in the number of friendships within the network.

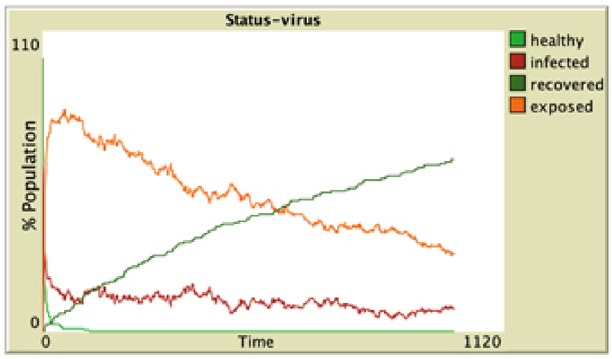

Fig. 3.1. Distribution of infection with *update-network-random*

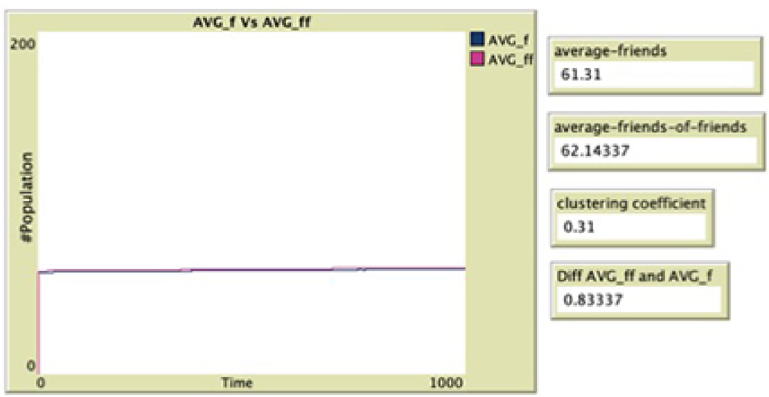

Fig. 3.2. AVG_f, AVG_ff and clustering coefficient

The same experiment was run again, using infection diffusion and *add or delete infected links*, setting the parameters *add-friendship-p* and *popular* to 0.7 and 59, respectively, the latter chosen to be close to the average number of friends.
From that experiment, it can be seen that the number of healthy decreases slowly until it remains constant, the exposed are few from the beginning and decrease steadily while the number of initially high infected decreases slowly while the number of recovered has an opposite trend to that of infected.
These trends are caused by the interaction between the agents that is obtained by using *add or delete infected links*, although the probability of adding new friendships is 70% however there remains that 30% probability of unfriending the infected and as much as they may continue to add links to healthy people this decreases the probability of infecting new individuals.
The clustering coefficient increased because the connections between the agents increased while, as seen for the previous experiment the values of AVG_f are always lower than AVG_ff but, in this particular case, the difference between the two values is consistent and the cause could be the removal of most of the connections to the infected, since this feature results in a greater inequality of the distribution of connections within the friendship network.
The results obtained can be observed below.

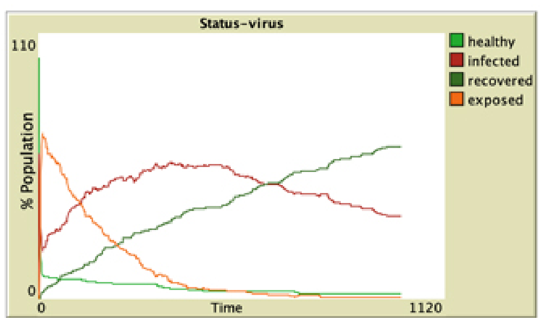

Fig. 3.3. Distribution of infection with *add or delete infected links*

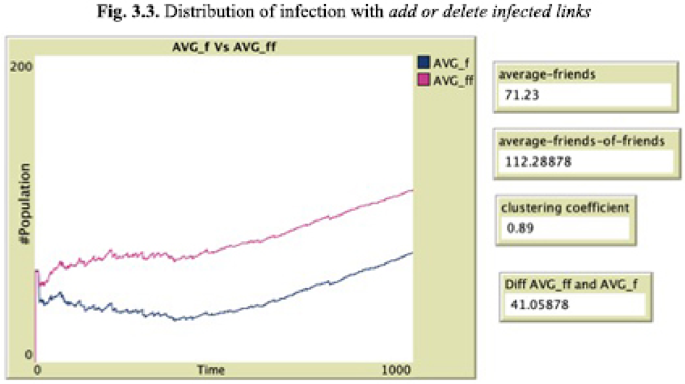

Fig. 3.4. AVG_f, AVG_ff and clustering coefficient

Game Theory

In the graph below, it can be observed the result obtained by using game theory and leaving the previously set parameters unchanged, such as *popular* set to 59. From these results, it can be observed that initially the number of no popular individuals is greater than the popular and that once game theory is started, new friendships will tend to be created between popular individuals up to and increase in this way the popularity of the individuals.
As for the clustering coefficient this increases because new connections are created making the network more connected while the value of AVG_f continues to be lower than that of AVG_ff.
The results obtained can be observed below.

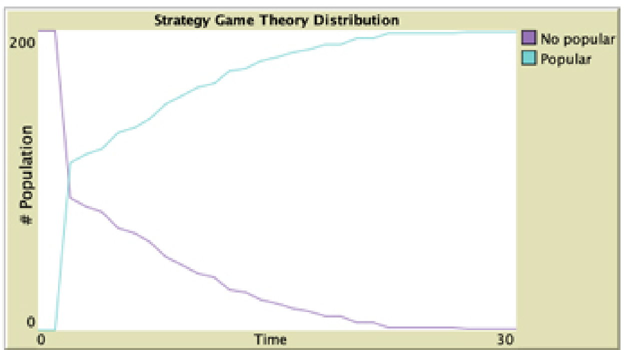

Fig. 3.5. Game theory distribution

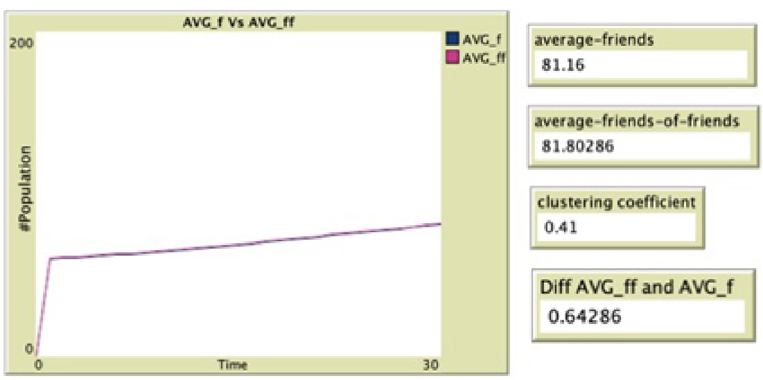

Fig. 3.6. AVG_f, AVG_ff and clustering coefficient

At last, an experiment was carried out by having the agents interact using all the models seen previously and leaving all the parameters unchanged.
In this case it can be seen that, as far as game theory is concerned, after a first phase similar to the previous one, the number of popular and non-popular individuals increase and decrease also decreasing their difference, this is because during the simulation also the network evolves going to eliminate friendships and adding others.
The values of AVG_f are always lower than AVG_ff the trend is the same as seen in the previous cases where we go to eliminate and avoid making connections with the infected, this leads to a greater inequality of the distribution of connections within the friendship network where some individuals have more friends than others, such as, the infected. This aspect can also be observed through the high clustering coefficient. The results obtained can be observed below.

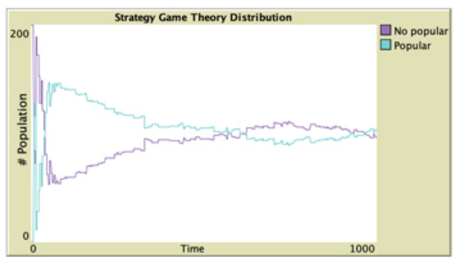

Fig. 3.7 Game theory distribution with *update-network-random* and *add or delete infected links*

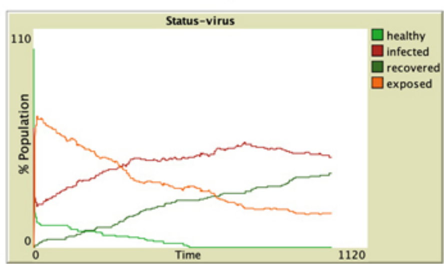

Fig. 3.8. Distribution of infection with *update-network-random* and *add or delete infected links*

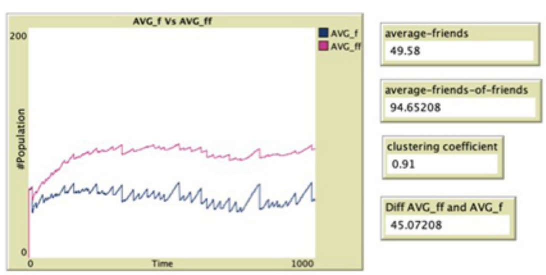

Fig. 3.9. AVG_f, AVG_ff and clustering coefficient

3.2 Second experiment

Diffusion infection

The results of the experiments below were obtained by creating a network with *num-links* equal 9857. The number of initial infected is 65 and the probability of *virus-spread-chace* and *recovery-chace* is 0.3 while *exposed-chance* is 0.5.
In Figure 3.10 and Figure 3.11 we note the results obtained by infection diffusion and *update-network-random*. Where it can be seen that the number of infected decreases slowly over time while the number of recovered increases.
The values of AVG_f and AVG_ff decreased slightly during the interactions while the clustering coefficient increased due to the high probability of the *drop-friendship-p* parameter set to 0.7. In this way although the number of friendships decreased, the probability that two individuals who are friends to a third agent are friends themselves increased.

In Figures 3.12 and 3.13, it can be observed the results obtained by leaving the parameters set as described above, but running the simulation using the spread of infection and add or delete infected link, setting the *add-friendship-p* parameter to 0.3. It can be seen that the number of infected has a gradual evolution while healthy decreases little in contrast to recovered that gradually increases.
The value of AVG_f decreased compared to the initial value while AVG_ff increased, also the difference between AVG_ff and AVG_f increased compared to the beginning this is because the probability to make friendship with healthy individuals is very low so during the simulation there will be more tendency to eliminate friendship with uninfected individuals. For the same reason, the clustering value is also higher than the initial state of the simulation where it was 0.50.

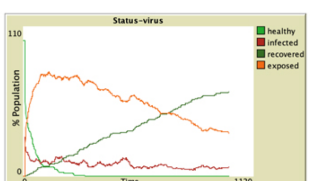

Fig. 3.10. Distribution of infection with *update-network-random*

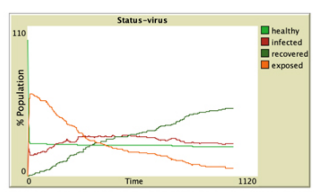

Fig. 3.12. Distribution of infection with *add or delete infected links*

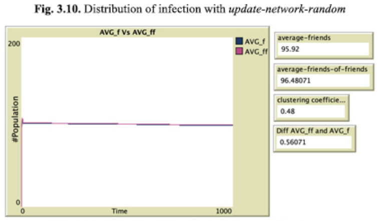

Fig. 3.11. AVG_f, AVG_ff and clustering coefficient

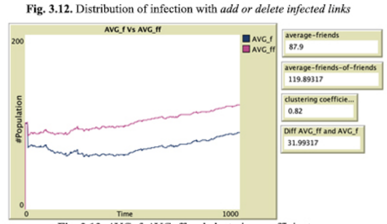

Fig. 3.13. AVG_f, AVG_ff and clustering coefficient

Game Theory

In Figures 3.14 and 3.15, it can be observed the performance of the simulation using only game theory, setting the *popular* parameter to 65, a value greater than both the average number of friends and the average number of friends of friends. It can be seen that the AVG_f and AVG_ff values increased steadily over time, always keeping the former value lower than the latter. In addition, the clustering coefficient has also increased by 0.09 from the beginning, and this may be caused by the constant increase of connections between popular individuals.
Lastly, in Figures 3.16 and 3.17 and 3.18, it is possible to observe the results obtained using game theory, infection spread, *update-network-random* and *add or delete infected links* simultaneously, keeping all previous parameters unchanged. From that experiment it can be seen that the diffusion of infection occurs in slow and gradual manners this is because although new links can be created with other individuals both infected and uninfected, there is also a high probability that friendships with infected people will be eliminated due to add or delete infected friendship. For the same reason, the number of popular and non-popular also have a slower trend. During this simulation, it can be seen that AVG_f and AVG_ff both increased and their difference is large and the clustering coefficient also has a high value. The last two features are obtained due to the addition of links only with the uninfected and to the elimination of friendships with the infected making the network more connected and clustered.

Fig. 3.14. Game theory distribution

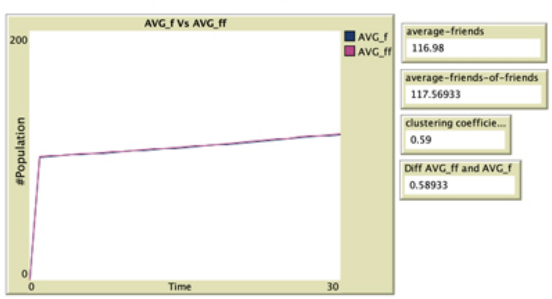

Fig. 3.15. AVG_f, AVG_ff and clustering coefficient

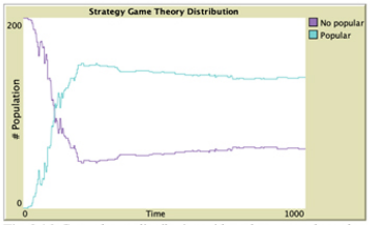

Fig. 3.16. Game theory distribution with *update-network-random* and *add or delete infected links*

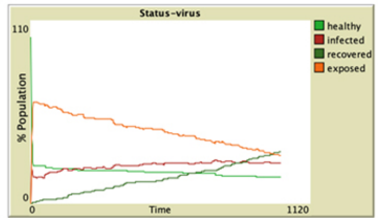

Fig. 3.17. Distribution of infection with *update-network-random* and *add or delete infected links*

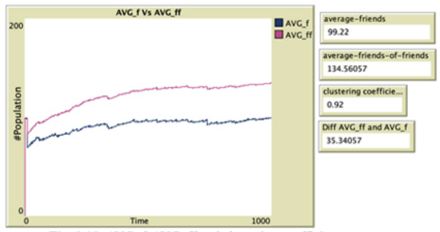

Fig. 3.18. AVG_f, AVG_ff and clustering coefficient

4 Conclusions

Although, the relationships and their interactions are difficult to predict due in part to multiple external factors that may influence these interactions, the results obtained from the model of infection diffusion and game theory seem to be quite realistic. In fact, very often people tend to avoid people who do not have a good reputation, excluding them and sidelining them to avoid being influenced by them.
In addition, the experiments conducted confirmed the hypothesis in which the average number of friends is always smaller than the average number of friends of friends and also it could be seen that as the difference between the two values increases, there is also an increase in the clustering coefficient. In fact, making the network of friendships more and more connected will tend to make the average number of friends of friends significantly larger than the number of friends, thus increasing the sense of inappropriateness in people who have few friends.
There is much work that could be done to be able to improve the model, such as adding additional payoff values to implement other models of interactions based on game theory.
However, the work done seems to be a good way to approach the problem on how a bad reputation can affect relationships and a person's popularity by going to decrease the number of friends and increase the number of friends of friends.

References

[1] Scott L. Feld. "Why Your Friends Have More Friends Than You Do". In: *American Journal of Sociology* 96.6 (1991), pp. 1464–1477. ISSN: 00029602, 15375390. URL: `http://www.jstor.org/stable/2781907`.