Complex Systems and Network Science
Exam Project Specification

Ozalp Babaoglu Nicolas Lazzari

2022/2023 Academic Year

1 Introduction

As part of your course requirement, you are to complete one of the projects described below, which must be carried out individually.

Submission of the project for evaluation must be done via email to the address: nicolas.lazzari2@studio.unibo.it.

The deadline for submission is 23:59:59 hours on 04 January, 2023. The email must have the subject field as *CSNS Project 2023* and must be sent from your University address (name.surname@studio.unibo.it).

You will receive a confirmation message within a few days of your submission. The email should contain an archive (in .zip or .tar.gz format) containing the following:

- The source code that was developed (either in NetLogo or PeerSim);
- A short paper, in PDF format, describing the model that was implemented, the experiments that were carried out using it, and a discussion explaining the results that were obtained.

Your full name, email address and student ID number (matricola) must be included in all of the source files, in the paper, and in the submission email that you send. The source code should be well documented and formatted, following good programming practices. The paper must be written in English, and should be structured like a technical paper, thus containing a title, abstract and bibliography. It is strongly suggested that you limit the length to 12 pages and that you follow the Springer format for Lecture Notes in Computer Science (LNCS). Templates are available for both Word [1] and LaTeX [2]. You can use any text processing system you prefer (even though LaTeX is suggested) to write the paper as long as you submit the result as a PDF file. The project must be done individually: no sharing of papers or source code is permitted. You are, of course, encouraged to discuss issues and solutions with fellow students or with the instructors.

[1] .doc template
[2] .tex template or overleaf template

2 Project specification

The project you will have to complete is the development of a simulation based on the paper published by Scott L. Feld on the American journal of sociology, Why Your Friends Have More Friends than You Do [3]. The paper describes an interesting situation based on real world data. The main theme addressed by the paper is that it is reasonable to suppose that individuals use the number of friends that their friends have as one basis of comparison to measure how adequate they are in their social circle. If individuals make this type of comparison it is likely that most of them will feel inadequate.

2.1 The problem

The basic logic can be described in very simple terms: if there are some people with many friendship ties and others with few, those with many ties show up disproportionately in sets of friends. For example those with 40 friends show up in each of 40 individual friendship networks and thus can make 40 people feel relatively deprived, while those with only one friend show up in only one friendship network and can make only that one person feel relatively advantaged.

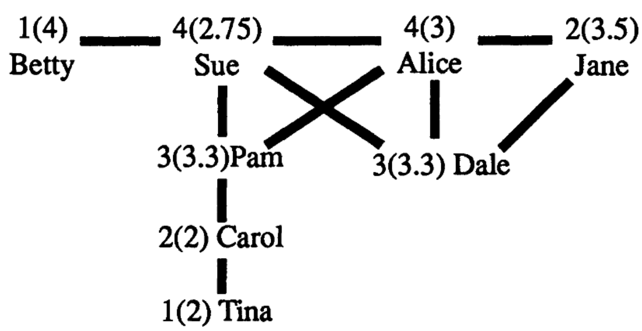
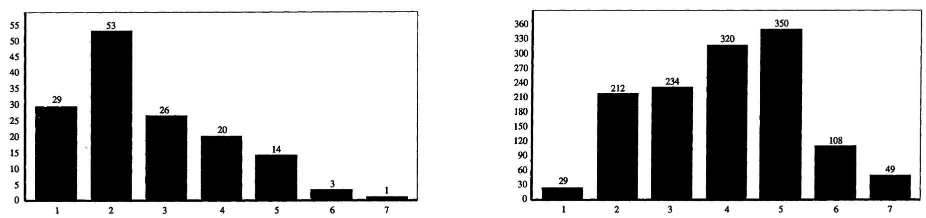


Figure 1: Friendship examples based on data from Marketville High School

Figure 1 shows an example extracted from The Adolescent Society [2], in which individuals from 12 different high schools were asked to name their friends. The number above each name represents the number of friends of each individual and the number in parentheses the mean number of friends of each friend. Only *Sue*, *Alice* and *Carol* have at least the same number of friends as the number they will compare to or more, while every other individual will feel inadequate when comparing herself to other individuals. The overall structure can be observed in the whole network as well.

In order to try and understand this phenomena we can begin by observing the distribution of number of friends and the distribution of number of friends of friends. The former is just the usual distribution of numbers of friends that we would examine in social science. In the example of figure 1 that would be a mean of 2.5 friends per individual. The distribution of friend of each friend, however, takes into account some individuals more than one time. For instance, if we take *Jane* as an example, then *Dale* and *Alice* contribute from her behalf to the number of friends' friends. When taking into account *Sue*, then along with *Pam* and *Betty*, *Dale* and *Alice* will be counted again. In fact each individual's friend contributes to the final friends' friends average as many times as she have friends.

Whenever each individual compares its number of friends with the average number of friends' friends the comparison is unfair: two different distributions are taken into account.



(a) Average number of friend distribution

(b) Average number of friends' friends distribution

Figure 2: Marketville High School distributions

In figure 2 the distributions are shown over all the students in Marketville High School. Figure 2a shows a typical avergae number of friends distribution, with a mean of 2.7. In figure 2b instead it's clear how the distribution of number of friends' friends, with mean 3.4, is skewed towards higher values. 74% of the students would feel inadequate if they used the average number of friends' friends as a comparative measure.

There is a simple relationship between the distributions in figure 2. The average number of friends of each individual is simply

$$AVG_f = \frac{\sum_{i \in I} x_i}{N} \tag{1}$$

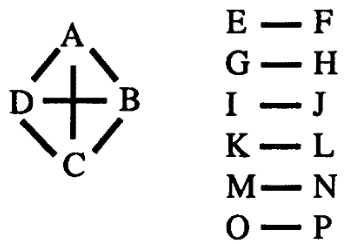
with x_i being the number of friends of the individual i, N the total number of individuals and I the set of individuals. On the other hand, the average number of friends' friends of an individual is

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

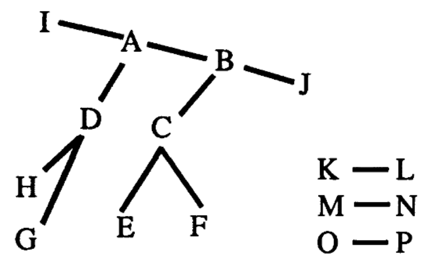
since an individual has x_i friends and contributes to each of its friends average as well. Equation 2 can be shown to be equal to

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

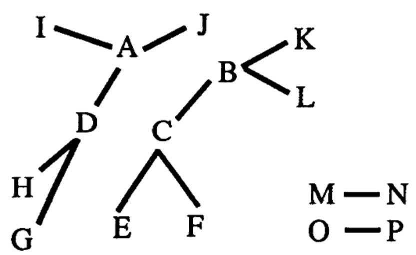
See [3] for a more detailed description on this can be derived. In the particular case in which $\sigma^2(x) = 0$ then the average number of friends' friends of each individual is simply the average number of friends $\mu(x)$ of the whole set of individuals. Depending on how friends are distributed among individuals, very different situations may arise.



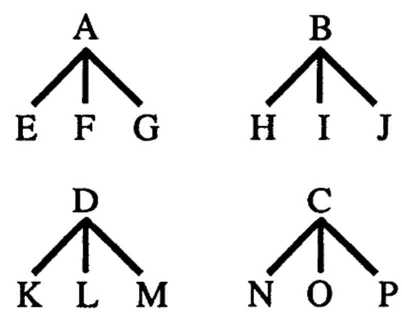
(a) Perfect positive correlation between $AVG_f = AVG_{ff} = 1.5$



(b) No correlation between $AVG_f = 1.5$ and $AVG_{ff} = 2$



(c) Negative correlation between $AVG_f = 1.5$ and $AVG_{ff} = 2.17$



(d) Perfect negative correlation between $AVG_f = 1.5$ and $AVG_{ff} = 2.5$

Figure 3: Network configurations impact on AVG_{ff} given AVG_f.

In figure 3 we can see the effect of the network configuration on the average number of friends' friends, given that the distribution of individuals friends is always the same: each individual $AVG_f = 1.5$. The impact of the network is indeed an important factor that plays an essential role in the distribution of AVG_{ff}. Social relationships are often tightly bounded with the concept of popularity. Individuals tend to choose friends that are more popular than they are, in the attempt to lift their social status. This prevents the creation of *healthy* network configurations, and as an outcome in cases the majority of individuals will have to compare themselves to highly unrepresentative samples from the population and feel inadequate.

2.2 The required model

You can see the requirement of the project as divided in two main fundamental blocks: building a model of the social network described in Section 2.1 and extending the implemented model to simulate social interactions between the individuals. If you wish to change the theme of the project, for instance taking into consideration the international economic relations between different states (friendship would be the commercial alliances between states and friends' friends the perceived economic power of a state) you're allowed to do so, as long as your extension is fundamentally in line with what is described here-after and any modification has been previously approved by email at the address in section 1.

2.2.1 Network model

The whole project revolves around the phenomena described in section 2.1. The environment of the simulation will be populated by N agents. Each agent represents a single individual that is linked to other agents. Each link represents a relationship. The initial structure of the network, i.e. the initial set of relationships, is the first modeling choice you have to take. Here you can choose any network configuration model, from the one that you've analyzed during lectures (e.g. Erdős-Rényi, Watts-Strogatz, Barabási-Albert) to other models that you think would better fit the model. See [4] for a brief survey on additional network models.

In your report you need to clearly state your modeling choice and the reason you've decided to do so. If you decide to implement a network structure that has already been explained during lecture (i.e. without applying any particular extension), you can avoid describing it in your report. You should, however, explain in detail how you expect each parameter to influence the overall simulation and, whenever you decide to use a fixed static value, motivate your choice. For example, if you decide to use the Erdős-Rényi model presented during lecture you should write on your report which values of the probability p you are going to try and why. You can test as many network configurations as you'd like. Try to test and analyse the experiments that you do as much as possible before adding additional features though. A simple model is often able to approximate many different configurations just by using the right combination of parameters.

2.2.2 Individual interactions

As introduced in section 2.1, relationships are often based on popularity and they might change over time. You need to model this aspect in your model. This means that, during the simulation, the network of your model will evolve into a different structure. The way this evolution is modeled and implemented are entirely up to you. Here are some different ideas on how to perform those changes, but you are encouraged in proposing your own:

- randomly change links. In this case your model is simply going to randomly sample links between agents and remove them. The sampling strategy is again your choice. For example, to simulate the differences between groups in social real-world networks, you can split your individuals in different clusters and sample with a different distribution from each cluster;

- treat relationships changes as an infection diffusion model. You can for instance identify a group of people as being infected by a *"social infection"*. Those individuals can spread the infection, and of course other individuals will try to avoid them. You can model many social phenomena using this considerations: think for instance of the influence that social media might have in the reputation of individuals and how that negative reputation spread among those who are close to that individual;

- model relationship changes using game theory. Here you are going to define an utility function for each agent. Agents are then going to play games with each other, in which they decide whether to remove (or respectively add) a link between them based on the outcome that their action would have on their utility function. You can see utility function, in the context of the problem of section 2.1, as the popularity of an individual. For additional examples based on game theory, refer to section 3;

- model relationships using a genetic algorithms. Similarly to the game theory approach, you can define a fitness function for each individual (e.g. popularity in the context of the description of section 2.1) and condition the decision of keeping, dropping or adding new links based on the individuals chromosomes.

Each of the proposed variations can be combined and/or extended as much as you like by picking any argument that has been addressed during lectures. Remember that models with too many parameters might quickly turn into chaotic models. If you can't understand the results of your model, simplifying some parts of it is a good rule of thumb to obtain simpler results.

2.3 General considerations

Regardless of your modeling choices, you should always try to add some noise to your model. For instance whenever you happen to remove a link, develop your model in a way that, by setting a parameter, there's a probability p_N that the removal won't take place. This will help you assess how robust your model is and if it is indeed able to model real-world scenarios, where random noise can't be ignored. As already mentioned, always keep in mind that it's very easy to build an unpredictable and unstable model. If you have any doubts on how to model some specific aspect send me an email. Make sure to insert *CSNS Project - 2023* in the object or I might miss it. Please don't send the model asking for help in debugging it. Instead provide a clear description of what you're trying to achieve and a clear description of where you are not sure on how to proceed.

The model you develop is not required to obtain optimal results. It is more important that you explain in detail what you expect from your model and how those expectation are (or aren't) fulfilled during the simulation. In both cases we expect you to explain how the parameters of your simulation change the final outcome and how the simulation performs under *"stressing"* conditions. You're not required to perform a rigorous analysis of the model (the scope of the project itself is to simulate situations that are otherwise difficult to approach analytically) however try to be as specific and precise as possible.

During your analysis you should at least analyse how the values of AVG_f and AVG_{ff} evolve under different initial conditions. If your model make use of additional parameters, describe the purpose of those parameters and which role they play in the simulation. If you decide to keep track of additional metrics, for instance the clustering coefficient of the network, try to propose an explanation of it.

3 Game Theory

To get an idea of how simple game theory approaches can be used to model a wide range of phenomena we will take a look a the paper *Coordination vs. voluntarism and enforcement in sustaining international environmental cooperation* by Scott Barrett[1]. The main theme is why and how cooperation between world-wide states can be modeled in the context of environmental actions. Two main events are taken as examples of successful international cooperation events: the eradication of smallpox achieved in 1980 and the protection of the stratospheric ozone layer, formalized in the Montreal Protocol in 1987, that has shown ever improving signs of recovery in recent years. While international cooperation succeed in those two events, it has mostly failed in every other instance it was proposed of. With the Paris Agreement of 2015 it is more important than ever to correctly analyze why international cooperation might succeed and fail, in order to develop an effective protocol in fighting the global climate emergency.

International cooperation on climate issues has traditionally been modeled as a cooperation game, in which each state either decided to negotiate national reductions in emissions or pledge to reduce their emissions voluntarily in the hope that a spirit of international cooperation arises. This approach has shown to be an over-simplification of the subject and Barrett proposes several new ways on how to model this interactions in a more complete way.

Public Goods Game Humans footprint on environmental catastrophes are sometimes modeled as a variation of the classic Prisoner Dilemma, extended up to N players, called *public goods game*. Each player represents a state that needs to perform a binary choice: either collaborate C or defect D. Suppose that each country is treated in an equal manner then all need to optimize the payoff function

$$\pi_i = (\overline{Y} - Y_i) + b(Y_i + Y_{-i}) \tag{4}$$

where $i \in States$, $0 < b < 1$ is the marginal benefit of each state contributing, $\overline{Y}$ represents the amount of money at disposal of a state, Y_i is the amount of money that a state is willing to contribute to the emergency and $Y{-}i$ is the sum of the amounts every other country is willing to contribute. It's easy to see that the global best possible result is obtained when every state contribute their maximum $\overline{Y}$ to the cause. However from the singular point of view of each state, it's better to contribute nothing and just benefit from other states contributions.

Treaty Game People cooperate more effectively with the aid of institutions. Barrett provides a deep and insightful description on how the cooperation between states might be more effective and reliable when an institution enforces democratically signed treaties. To model this aspect the previous public goods game is extended to a 2-rounds game. In the first round of the game each state decides whether to sign the treaty or abstain from it. In the second round the public goods game is played. In this instance, however, the decision of all those states that signed the treaty will be to collaborate and every state that decided to abstain will independently decide to collaborate or not. In particular each state that signed the treaty decides a fixed amount of money they are all going to contribute. When the treaty game is played as a one-shot game, collective cooperation is not easily achievable. Real world scenarios are better represented by a repeat version of this game, in which each state has the choice to change its first game decision.

Punishments for non-collaboration It's reasonable to think that, when a state is contributing money to fighting climate issues, one important condition is that the outcome needs to be fair for each state. That is, the marginal gain of those who collaborate needs to be higher than the passive gain of those that decides to defect the collaboration. This can be modeled as a reduction of the marginal benefit b for those states that decide to avoid collaborating with other states.

Catastrophe Avoidance game In the climate change game, one could see the result as a simple binary outcome: either fixing the issues or passing the "tipping point" resulting in a catastrophic loss. To model this kind of behaviour it is sufficient to update slightly the public goods game. A new variable $\overline{\overline{Y}}$ is added to the model, which represents the minimum amount of money needed to win the climate fight game. Each players payoff function can hence be updated to

$$\pi_i = (\overline{Y} - Y_i) + b(Y_i + Y_{-i}) - X \tag{5}$$

for $Y_i + Y_{-1} < \overline{\overline{Y}}$ where X is the loss that each state will sustain if the fight is not won and

$$\pi_i = (\overline{Y} - Y_i) + b(Y_i + Y_{-i}) \tag{6}$$

for $Y_i + Y_{-1} \geq \overline{\overline{Y}}$.

References

[1] Scott Barrett. Coordination vs. voluntarism and enforcement in sustaining international environmental cooperation. *Proceedings of the National Academy of Sciences*, 113(51):14515–14522, 2016.

[2] James S Coleman. The adolescent society. 1961.

[3] Scott L Feld. Why your friends have more friends than you do. *American journal of sociology*, 96(6):1464–1477, 1991.

[4] Anna Goldenberg, Alice X Zheng, Stephen E Fienberg, Edoardo M Airoldi, et al. A survey of statistical network models. *Foundations and Trends® in Machine Learning*, 2(2):129–233, 2010.