TÉCNICO LISBOA

Engineering Trust in Complex Networks

Manuel Tavares de Sousa Rosa Galamba

Thesis to obtain the Master of Science Degree in

Information Systems and Computer Engineering

Supervisors: Dr. Fernando Pedro Pascoal dos Santos
Prof. Francisco João Duarte Cordeiro Correia dos Santos

Examination Committee

Chairperson: Prof. João António Madeiras Pereira
Supervisor: Dr. Fernando Pedro Pascoal dos Santos
Member of the Committee: Dr. Andreia Sofia Monteiro Teixeira

June 2020

Acknowledgments

I would like to express my sincere gratitude to everyone who helped me make this project possible. In no particular order, I would like to thank: my dissertation supervisors, Prof. Francisco Santos and Dr. Fernando Santos, for their guidance, support, sharing of knowledge and constructive criticism. My parents, my family, Marta and my friends for always being present during the good times, and for lifting me up during the more difficult times, for all their advice and support.

Abstract

In a world where every day more economic transactions are done via the internet, Trust and Trustworthiness are pivotal for many services to work properly. To study both of these elements, through game theory, it is common to use the Trust Game. Even though game theory dictates that in one-shot interactions, namely by means of the game's unique Nash equilibrium, investors should not trust the trustees nor should these be trustworthy, behavioral data from several experiments shows the opposite. Hence, there is the question of how trust can be stabilized in the original Trust Game in order for it to capture what happens in reality. Several studies emerged, addressing versions of the game, that consider reputation or are played in social networks, the effects of combining both of these components, however, are not clear. In this work, we propose a new model, consisting of a Trust Game version with reputation, using Evolutionary Game Theory as a framework, played in both finite unstructured populations and in static social networks, where we introduce other variations with the objective of increasing Trust and Trustworthiness in the population. We conclude that taking into account players' reputation has a positive effect. When played in a Social Network, the introduction of network based role and strategy assignment, namely based on individuals' degree in the Network, may yield a considerable increase of Trust and Trustworthiness. The most successful variations were when considering the more connected individuals as Investors and the introduction of pathological players in the population.

Keywords

Trust, Evolutionary Game Theory, Social Networks, Reputation, Game Theory

Resumo

Num mundo em que são feitas cada vez mais transações económicas pela Internet, a Confiança e a Confiabilidade são essenciais para que muitos serviços funcionem corretamente. Para estudar estes dois elementos, utilizando teoria dos jogos, é comum o uso do *Trust Game*. Embora a teoria dos jogos dite que, nas interações de uma só jogada, considerando o único equilíbrio de Nash do jogo, os investidores não devem confiar nos receptores, nem devem estes ser confiáveis, dados comportamentais de várias experiências mostram o contrário. Existe, portanto, a questão de como é que a confiança pode ser estabilizada no *Trust Game* original para capturar o que acontece na realidade. Embora existam vários estudos sobre versões do jogo que consideram reputação ou são jogadas em redes sociais, os efeitos destas componentes em conjunto não são claros. Neste trabalho, propomos um novo modelo, que consiste numa versão do *Trust Game* com reputação, usando Teoria dos Jogos Evolutiva, jogada tanto em populações finitas não estruturadas como em Redes Sociais, onde introduzimos outras variações com o objetivo de aumentar a Confiança e a Confiabilidade. Concluímos que ter em conta a reputação dos jogadores tem um efeito positivo. Quando jogado numa rede social, a introdução de papeis e estratégias atribuídas com base na rede, nomeadamente os graus dos indivíduos na rede, tem por vezes como consequência um aumento considerável de Confiança e Confiabilidade. As variações mais bem-sucedidas ocorreram quando os indivíduos de maior grau foram considerados investidores e com a introdução de jogadores patológicos na população.

Palavras Chave

Confiança, Teoria dos Jogos Evolutiva, Redes Sociais, Reputação, Teoria dos Jogos

Contents

List of Figures

List of Tables

Acronyms

SF	Scale-free
ER	Erdös-Rényi
EGT	Evolutionary Game Theory
ESS	Evolutionary Stable Strategy
PROP	Proportional Imitation
UI	Unconditional Imitation
UI-VM	Hybridization of Unconditional Imitation and Voter Model
MO	Moran Process
TC	Trustworthy Consumer
UC	Untrustworthy Consumer
TP	Trustworthy Provider
UP	Untrustworthy Provider

1

Introduction

Contents

Trust and trustworthiness are fundamental to a successful society. In today's world, many social and economic transactions occur through the internet between people that will never actually meet in real life. In order for these transactions to work, individuals must expect that their partner in the transaction will not behave opportunistically in an attempt to maximize their own payoff, regardless of the fact that their decisions could possibly cause prejudice to their counterpart. Such trust, however, is not easy to explain or sustain.

In economics, many times, one of the main assumptions is that individuals will act in their own self-interest since this way they maximize their payoff. Therefore, in individual choice settings, an individual choosing an action that deviates from his self-interest is considered irrational (assuming that individuals act following a utility function that only accounts for their own gains). In group settings, however, there are situations where acting in self-interest will make all the individuals worse off [1]. In an attempt to understand human behaviour in this last situation, Berg et al. [1] developed an experimental setting named the Investment (or Trust) Game.

The Trust Game, in brief, is an interaction between an investor and a trustee. The investor initially has a certain amount of money that he can either keep or transfer to the trustee. This value is multiplied by a factor $b > 1$ before reaching the trustee which will then decide how much to return to the investor [1].

Based on the mathematical models that are used to study this kind of interactions, namely Game Theory, one would assume that people playing the Trust Game would act to maximize their own payoff, particularly when considering that the unique Nash Equilibrium [7] for this game is for the investors to transfer zero money [1]. Behavioral data resultant from real experiments with this game, however, reveals that investors do make transfers and trustees return considerable amounts, *e.g.* [1, 8].

Because traditional Game Theoretical Models fail to predict what actually happens when the Trust Game is played, there was still the question of how high investment preferences arose, how they are maintained, as well as how would they change when interactions occur in populations, as in a real life scenario, instead of between just two players. With the objective of answering these questions several studies emerged where the authors add some complexity to their model, these are detailed in chapter 3.

The focus of this Thesis is the study of the Trust Game, namely through the use of computer simulations, as well as several variants of the same. To do this we propose our own model where we combine different types of complexity, similar to the ones in other studies, and add some new. In particular, as detailed in the objectives section, we are going to study the effects of introducing reputation in the game, firstly in well-mixed populations and secondly in structured populations, specifically in a scale-free network. The introduction of reputation is justified by the fact that, when comparing with the behavioral data, we assume that the people partaking in the experiments developed their strategies in broader contexts. Manapat et al. [3] points that in the context of daily life, many times, investors have information about the trustees due to the existence of reputation systems and that this favors the promotion of cooperation [9].

The usage of networks comes from the fact that in real life scenarios populations have a finite number of individuals and a structure. By having the players represented as nodes of a network with connections only to their direct neighbours, it allows us to study some properties exclusive to networks, namely having a player role, reputation or even strategy dependent on the number of connections he as in the network.

The results of the simulations of our model, as detailed in chapter 5, show that the introduction of both reputation and structure to the population playing the Trust Game, particularly for certain variations of the game that use some of the networks' properties, lead to an increase in the promotion of both trust and trustworthiness.

1.1 Objectives

This report will initially focus on studies on different variants of the Trust Game, mainly through the use of computer simulations, realized by various people in an attempt to understand the behaviour of individuals regarding trust and trustworthiness, specifically how does this behavior arise and how does it maintain.

The objective of this work, however, is to advance our understanding of this topic through the use of our own model. As mentioned above, real-world sharing economy transactions usually consider some kind of reputation. The model we propose consists of applying a reputation system to: firstly finite unstructured populations; secondly, a networked version of the Trust Game, namely with a static scale-free network, using evolutionary game theory as a framework.

Many times, considering infinite populations is more convenient from a mathematical point of view, however, real world populations are finite and considering these instead introduces considerable changes sometimes [10]. For the version using unstructured populations, our goal is to verify if the results regarding infinite populations in [2] are extendable to finite populations. To study this we first consider the case where there are two populations, one of each role (investors and trustees) and every individual from either of the populations interacts with all of the individuals from the other population. Additionally, we run the same simulations for the symmetric version of the game, *i.e.* only one population where every individual plays as both roles while still interacting with all of the other members of the population.

Furthermore, for the networked version, we explore three different scenarios:

1. Asymmetric role assignment (λ model)
2. Diversity in the reputation
3. Hybrid societies with pathological players

Regarding asymmetric role assignment, we consider a λ parameter that controls how much a player's role (either investor or trustee) depends on his degree in the network. The main goal here is to study network-based role assignment, *e.g.* where highly connected individuals may also be in a better position to play as investors. Previously, degree-based role assignment was shown to affect fairness in Ultimatum Games [11]. However, it remains unclear whether the same occurs in Trust Games.

The second scenario consists of assigning different reputation values to players according to their degree in the network. We do this by forcing players with a higher degree to have a higher reputation value. To compare results we also do some experiments where we assign the reputation values of each player using a random distribution. For this scenario, every player plays as both roles.

Finally, the third scenario consists in having pathological players, *i.e.* a group of players that, regardless of the time step of the simulation and the player they are interacting with, will always act cooperatively during the experiment. Our main objective with this scenario is to see whether having the pathological players assignment dependent on the players' degree in the network will have an impact on the promotion of trust and trustworthiness. To study this we ran two kinds of experiments, one in which we assign the pathological players according to a random distribution and another one where we assign them according to the players' degree, namely by defining a threshold for the degree, above which those players will always cooperate. Previous studies show that small fractions of pathological players significantly affect cooperation and fairness dynamics [12–14]. We will extend that study, for the first time, to Trust and Trustworthiness in the case of Trust Games.

1.2 Outline

In this chapter, firstly we introduce the focus of this Thesis as well as our motivation and highlight some works that contributed to this, and secondly describe the objectives of our own study. In chapter 2 some background theory is provided that we think is helpful to better understand the following chapters. The related work we found most important is detailed in chapter 3, followed by our proposed model in chapter 4 which is divided into: firstly, some general information about how the Trust Game will be played and, secondly, some details regarding the two types of population structure considered. In chapter 5 we follow a similar structure, presenting firstly the methods, secondly the results of our computational simulations with the game and ending with a discussion of these results. Lastly, we close this document with some conclusive notes, a summary of our contributions and introduce some potential future work in this area.

2

Background Theory

Contents

In order to create our simulations in this field, as well as to properly comprehend the related works analyzed in the next chapter, we need to understand a variety of concepts related to game theory, of which we present the most relevant in this chapter. These concepts will allow to, not only better design these simulations, but also more accurately interpret the results obtained from them.

2.1 Game Theory

The purpose of game theory is to help us understand strategic interactions between decision-makers through the use of mathematical models. These decision-makers are often referred to as players, who act rationally according to a set of rules, hence the usage of the word "game". The scope of game theory is however much larger, varying from economic, like the Trust Game, to biological or even political phenomena [15].

These mathematical models are composed of three elements: a set of players, a set of actions that are available for each player, and a specification of each player's preferences. Every player knows the set of actions for all the players, including himself, and the resultant payoff from all the combinations between their actions and the other players'. This is often represented by a payoff matrix or a decision tree. Based on this information, for every interaction, players must select the actions - or sequences of actions - that most likely will maximize their payoffs.

2.1.1 Nash Equilibrium

A Nash Equilibrium [7] corresponds to a set of actions with the property that, if every player adheres to this set, no individual player will do better by choosing a different one.

In a formal definition [15], let a_i be the strategy profile of player i and a_{-i} be the strategy profile of all the other players, then a^* is a Nash equilibrium if:

$$\forall i, a_i : u_i(a_i^*, a_{-i}^*) \geq u_i(a_i, a_{-i}^*) \tag{2.1}$$

and is a strict Nash equilibrium if:

$$\forall i, a_i : u_i(a_i^*, a_{-i}^*) > u_i(a_i, a_{-i}^*) \tag{2.2}$$

where u_i is the payoff function for player i

2.2 Evolutionary Game Theory

Evolutionary game theory (EGT) was first introduced in [16] and consists in the application of game theory to populations. It originated in a biological context and it comes from the realization that the average payoff, here named fitness, of particular phenotypes, meaning the observable characteristics or traits of an individual, depends on their frequencies in the population [17]. In recent times, however, EGT has been of interest in other fields like economics and sociology. Here instead of measuring the fitness of a phenotype, we measure the fitness of a strategy in terms of how successful (*e.g.* economically) it is.

Evolution, in this context, works by selecting the individuals that perform better than average, modeling Darwinian competition. In EGT, however, instead of the less apt individuals dying, evolution occurs due to social learning. Individuals with better fitness are imitated by the others, according to a certain update rule.

2.2.1 Evolutionarily Stable Strategy

As stated before, with EGT we are applying game theory to populations, so instead of Nash equilibria, one can now consider a Evolutionarily Stable Strategies (ESS). Similarly to the Nash equilibrium, an ESS is a strategy such that, if adopted by all the individuals in the population, then no mutant strategy can invade the population [17]. ESS can, this way, be considered a refinement of a Nash Equilibrium: all strategies constituting an ESS form a Nash Equilibrium but the inverse is not necessarily true. In a formal definition [17], let the fitness of an individual with strategy A in a population where all the other individuals have strategy B, be written as $W(A, B)$. Then for strategy I to be an ESS we must have:

$$\forall I \neq J : W(I, I) > W(J, I) \tag{2.3}$$

or

$$\forall I \neq J : W(I, I) = W(J, I) \; and \; W(I, J) > W(J, J) \tag{2.4}$$

2.2.2 Well-mixed populations

Evolutionary games have traditionally dealt with infinite unstructured populations (well-mixed populations), in which each agent interacts with all other agents with equal probability. This setup can be conveniently described through the so-called replicator equation [2, 18], a deterministic equation which allows the study of fitness-based evolution in time. This equation may define both genetic evolution or a process of social learning in which, in the first case, individuals with higher fitness will reproduce more or, in the later, individuals with higher fitness will tend to be imitated more often. In any case, strategies that do better than average will grow, whereas those that do worse than average will diminish. As usual, fitness is here defined as the average return each agent gets from interacting with all the other members of the population.

Following Bohnet and Zeckhauser [19] and Berg et al. [1] we can define a payoff matrix for a very simple version of the Trust Game with two populations and four possible strategies. The first player, the investor, may choose between: sure-thing (S) and Trust (T), where sure-thing means not trusting and keeping the endowment. The second player, trustee, must adopt one of two strategies: Betray (B) or Reciprocate (R).

Initially, investors have an endowment of 1 and only three variables are considered:

p : Investor offer, $0 < p \leq 1$;

q : Trustee payback q (fraction of the multiplied value, mp), $0 \leq q \leq 1$;

m : multiplicative factor, $m > 1$;

Table 2.1 shows the resultant payoff matrix.

Table 2.1: Payoff Matrix for a simple version of the Trust Game

	B	**R**
S	$(1, 0)$	$(1, 0)$
T	$(1 - p, pm)$	$(1 - p + pqm, pm(1 - q))$

Note that, following the concepts introduced above, only the pair of strategies (S,B) constitutes a Nash Equilibrium - as long as $qm > 1$ and $q > 0$. As for the application of EGT, we shall assume that the evolution of the frequencies of strategies in both populations occurs according to the replicator equation. Following Sigmund [2] we can derive a system of two equations, describing the evolution of both the investors and the trustees' strategies.

Let us consider $e = (e_1, ...e_n)$ as the vector with all the possible strategies, and x as the vector that gives us the state of the population. This means each value of $x = (x_1, ..., x_n)$ represents the fraction of the population that uses that strategy, *e.g.* x_i corresponds to the fraction of the population using the pure strategy e_i. Assuming that populations can evolve then we must also assume that the frequencies x_1 change through time. In this manner we can say that $x(t)$ depends on time and that $\dot{x}_i(t)$ is the velocity with which x_i changes.

Sigmund in [2] states that the replicator equation holds if the growth rate of a strategy's frequency corresponds to the difference between its payoff and the average payoff in the population:

$$\dot{x}_i = x_i[f_i(x) - \varphi(x)] \quad (2.5)$$

$f_i(x)$ corresponds to the expected payoff of playing with strategy e_i (where a_{ij} corresponds to the payoff of playing as e_i against a player with strategy e_j)

$$f_i(x) = \sum_j a_{ij} \cdot x_j \tag{2.6}$$

and $\varphi(x)$ corresponds to the average payoff in the population, given by

$$\varphi(x) = \sum_i x_i \cdot f_i(x) \tag{2.7}$$

Let us consider x and y as the scalar numbers representing fractions of the population of investors and the population of trustees respectively and $p \cdot q \cdot m = \theta$. If x corresponds to the fraction of the investors' population with a cooperative (T) strategy (and $1 - x$ as the fraction with a defective (S) strategy) and y corresponds to the fraction of cooperative (R) players in the trustees' population (with $1 - y$ being the fraction of defective (B) trustees), then we have:

$$\begin{aligned} f_T(x) &= y\theta - p + 1 \\ \varphi(x) &= x(y\theta - p) + 1 \\ \dot{x} &= x(1-x)(y\theta - p) \end{aligned}$$

and

$$\begin{aligned} f_R(y) &= x(p \cdot m - \theta) \\ \varphi(y) &= x(p \cdot m - y\theta) \\ \dot{y} &= y(1-y)(-x\theta) \end{aligned}$$

From these equations we conclude that $\dot{y}$ will always be negative or 0, therefore y goes to 0. As a result $y\theta - p$ tends to be negative and x goes to 0, leading to no trust and no reciprocity. Fig. 2.1 shows the dynamics of evolution for this version of the Trust Game with $p = 1$, $q = 0.5$ and $m = 3$ (as mentioned before, given that $qm > 0$ we have that the pair of strategies (S, B) is the unique Nash equilibrium composed by pure strategies).

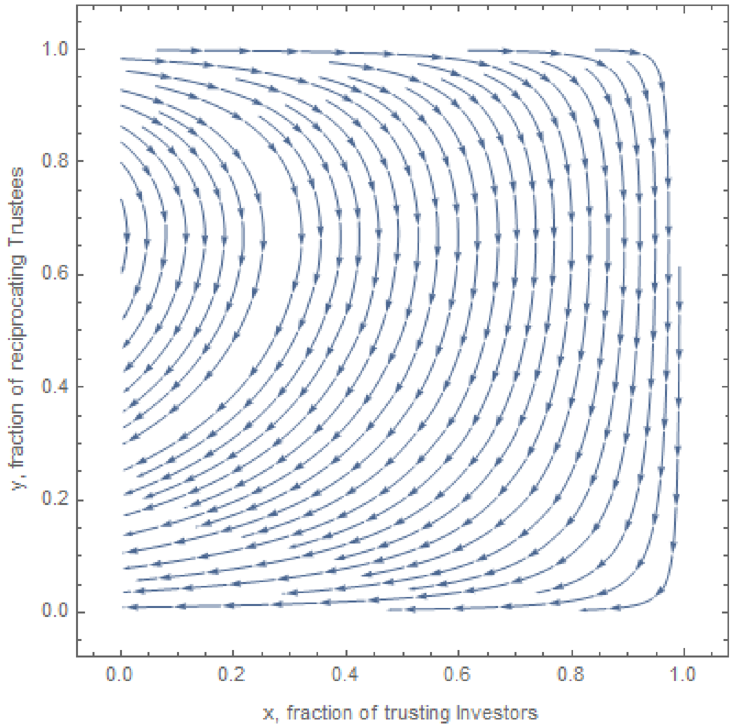

Figure 2.1: The replicator dynamics for the simple version of the Trust Game.

Following a similar idea to the Trust Game with reputation in [2], but still considering a simpler version of the Game, we can introduce a reputation $0 \leq r \leq 1$ to the previous payoff matrix (2.1) regarding the trustworthy trustees, *i.e.* r corresponds to the probability that a trustworthy trustee becomes known as such. The resultant payoff matrix can be seen in Table 2.2.

Table 2.2: Payoff Matrix for simple version of the Trust Game with Reputation

	B	**R**
S	$(1, 0)$	$(1 + rp(qm - 1), rpm(1 - q))$
T	$(1 - p, pm)$	$(1 + p(qm - 1), pm(1 - q))$

Once more, using equations 2.5, 2.6 and 2.7 we get

$$f_T(x) = 1 + y(-r \cdot p + r \cdot \theta)$$
$$\varphi(x) = yr(x\theta + xp + \theta - p)$$
$$\dot{x} = x[xy(-r \cdot p - r \cdot \theta - \theta) + y(-r \cdot \theta + r \cdot p + \theta) + xp - p]$$

and

$$f_R(y) = x(p \cdot m - r \cdot p \cdot m - \theta + r \cdot \theta) + r \cdot p \cdot m - r \cdot \theta$$
$$\varphi(y) = yr[-x(p \cdot m) + p \cdot m + r \cdot \theta - \theta] - yx\theta + x(p \cdot m)$$
$$\dot{y} = y[xy(r \cdot p \cdot m - r \cdot \theta + \theta) + y(-r \cdot p \cdot m + r \cdot \theta) + x(-r \cdot p \cdot m + r \cdot \theta - \theta) + r \cdot p \cdot m - r \cdot \theta]$$

Fig. 2.2 shows the dynamics of evolution for this version with $p = 1$, $q = 0.5$, $m = 3$ and $r = 0.5$ (from the payoff matrix in Table 2.2 with these values and eq. 2.1 we can conclude that there is no Nash equilibrium in pure strategies as long as $r < 1$).

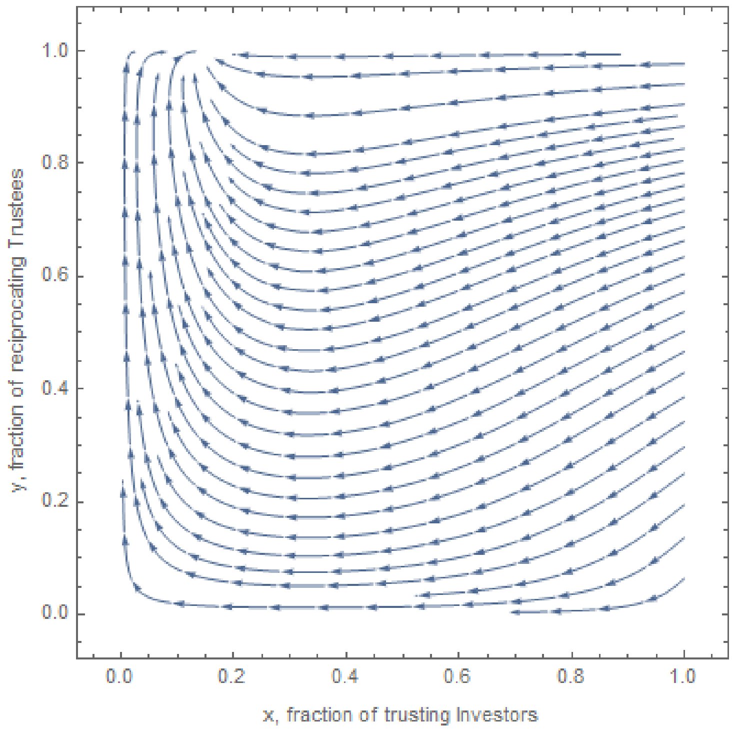

Figure 2.2: The replicator dynamics for the simple version of the Trust Game with Reputation.

2.2.3 Games on Graphs

Although the use of infinite unstructured populations may be more convenient from a mathematical point of view, in the sense that the replicator equation can be used to describe the dynamics of the populations, in real-world situations, populations are finite and individuals are constrained to interact with (and imitate) a subset of the population, an idea conveniently defined as a network: each agent is represented by a node that is constrained to play solely with its closest neighbours. The impact of topological constraints is known to induce profound evolutionary effects, as demonstrated experimentally in the study of the evolution of different strains of *Escherichia coli* (Kerr et al. [20]). In social settings, computational and mathematical models have also shown that cooperation is favoured on spatially structured populations [21]. This result has been recently demonstrated experimentally with humans [22].

In most social settings, and contrarily to spatially unstructured populations where all individuals (homogeneously) interact with the same number partners, some individuals engage in more interactions than others which, as a result, may potentially create conditions for a broad distribution of fitness values. Such heterogeneous scenarios often comprise a small number of nodes with many interaction links, called hubs, connecting the majority of nodes that contain fewer neighbours [23].

In this Thesis, we analyze both homogeneous and heterogeneous populations. For the latter, we adopt a paradigmatic example of such interaction structures: scale-free networks [24].

A network, also called an undirected graph, consists of a pair $G = (V, E)$, where V is a set of vertices, also named nodes, and E is a set of edges, *i.e.* the existing links between the network nodes. When two nodes are connected by an edge we consider them neighbours in the network.

Before defining scale-free networks we must first introduce the concept of degree distribution. When considering a network, the degree of a node corresponds to the number of connections it has with all the other nodes in the network. Consequently, the degree distribution of a network is the distribution of these degrees over the whole network. An SF network is a network whose degree distribution follows a power law for large k, *i.e.* $P(k) \sim k^{-\gamma}$, where $P(k)$ is the fraction of nodes in the network with degree k and γ is the exponent of that specific power law. In order to generate scale-free networks, the Barabási-Albert model can be used. This algorithm is detailed in the Networked Trust Games section of the related work chapter of this report (chapter 3.

For the case of simple one-shot 2-player games cooperation as the prisoner's dilemma, scale-free interaction structures were shown to help cooperation to thrive [25] when compared with homogeneous interaction structures, as highly-connected nodes are promptly taken over by cooperators who can then influence the whole community into cooperating. This enhancement is grounded on the diverse nature of real interactions. However, there are still a reduced number of studies on the impact of such structures on the evolution of trust, a question aimed by this Thesis.

3

Related Work

Contents

The objective of this work is to study the importance of trust, reciprocity, and reputation in the context of money transactions, or, generally, situations that require trusting another person or entity in order to achieve a payoff maximizing outcome. For that, we will use networks and evolutionary game theory to simulate the interactions between players. We believe that these three factors (trust, reciprocity, and reputation) play a very important role in a substantial number of investments (and even sales) nowadays, namely when these are made through the internet and therefore causing the usual face-to-face component to be nonexistent.

To do this we will use as a starting point the Trust Game (also called investment game) suggested by Berg et al. [1] in 1995. Although this study does not take into consideration reputation, it is one of the earliest experiments in the field to use a simple game interaction to systematically study the dilemma of trust and reciprocity.

The experiment consists of a group of subjects that are placed in a room (room A) and receive an initial amount of money. Each subject in room A must then choose how much of this initial endowment to send (they can opt to keep all the money) to another anonymous individual located in a second room (room B), knowing that the amount they send will be tripled by the time it gets to room B. After receiving the money, each subject in room B then must decide on how much money to send back to room A and how much money to keep.

The main reason we found this experiment of interest and consequently made us explore more of this field is the fact that assuming that all the subjects were rational, it was expected that the subjects from room A would not send any money to room B since this is the predicted and unique Nash equilibrium as it is shown in chapter 4 and in [1], in the trust game section of [2], as well as in many other studies in very similar games like the peasant-dictator game for the discrete case [26] or even for the N-player version of the game [27]. Furthermore, if subjects from room A actually did send any money, one would think that none of the subjects from room B would send money back. The results from the experiment in Fig. 3.1, however, show the complete opposite of this along with other behavioral experiments with the trust game [8, 19, 28–30].

In [1] the authors divide the experiment into two sessions, one with no history treatment in which subjects were not given any information about prior similar experiences, and one with a social history treatment in which new subjects were given the results of the first experiment before ”playing” the game. Although this provides some interesting results, namely that the average amount returned by subjects in room B increases with a social history treatment, we wanted to see the effect it would have if subjects had prior information about the other subjects with whom they are actually ”playing” the game, which would represent, in a sense, their reputation.

As it was previously said, the work presented in [1] was only a starting point which motivated many other works with the Trust Game. We will now present and analyze the results and published studies

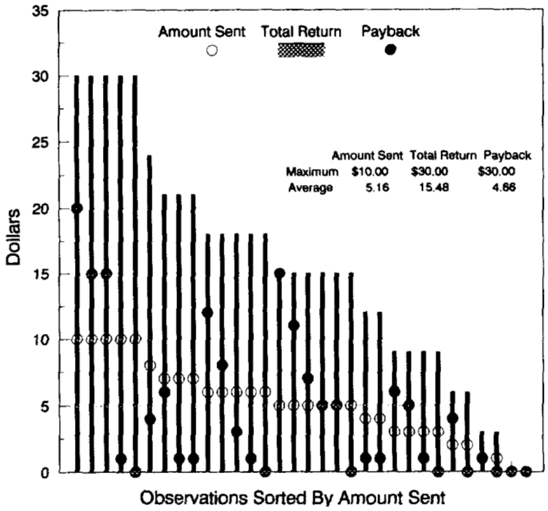

Figure 3.1: Trust experiment results with no history provided to the subjects, from [1]

that we found most important regarding Trust Games, reputation, and evolutionary game theory, and that will help us have a solid foundation to explore these three concepts together in our own study. While analyzing these studies, we will do some comparisons to our own model (which is detailed in chapter 4). When mentioning other experiments, subjects from room A, in this report, will be treated as Investors, and subjects from room B as Trustees.

3.1 Reputation in unstructured population Trust Games

One of the first studies to introduce reputation to Trust Games was written by Sigmund in [2]. This analysis considered, however, infinite populations, an assumption that we relax in this Thesis. The author firstly introduces a similar version to the Trust Game proposed in [1], yielding the same results predicted by the Nash Equilibrium, after the initial state the population evolves so that all the players become defectors (*i.e.* refuse to offer, or return, anything to the other player). Later, another version was introduced considering reputation, showing its positive effects on trust and trustworthiness. The evolutionary dynamics of this second version can be seen in Fig. 3.2 from [2], where the x axis corresponds to all the possibilities for the initial fraction of defective investors, y the fraction of defective trustees, and the arrows represent the direction of evolution from all the possible states. This work is particularly important to our study since we consider identical payoff matrices and the same reputation system.

Another important work regarding reputation in unstructured populations, although this time with finite populations, was done by Manapat et al. [3]. In their simulations, through the use of an unstructured population of investors and trustees, *i.e.* every investor interacts with all the trustees, the authors simulate the behaviour of the agents in a Trust Game where the Investors have, sometimes, information

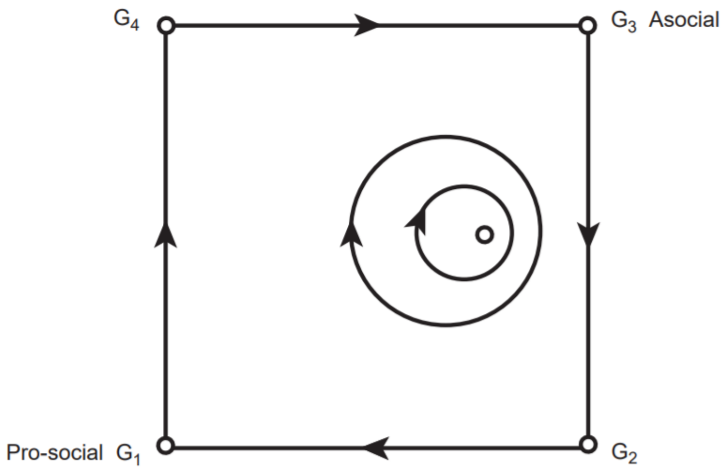

Figure 3.2: The replicator dynamics for the Trust Game, when considering that trustees have a reputation. From [2]

about the Trustees. For each interaction, randomly picking a pair of an investor and a trustee, the investor knows, with a probability q, the exact fraction r of the amount the trustee will return, q acts, this way, as a measure of the availability of information. An investor will always transfer money if $r > 1/b$, with $b > 1$ being the factor by which the stake is multiplied if the transfer is made.

We believe that this resembles our own model in the sense that the r rate of returning the money acts basically as a reputation system. The main difference lies in the fact that, because investors always act rationally, as long as $r > 1/b$ they will make the transfer. In our reputation system, however, defective investors will only transfer an amount proportional to the trustee's reputation.

It should also be noted the fact that in [3] there is a probability $1 - q$ (with $q < 1$) that the investor does not have any information on the trustee, which does not happen in our reputation system since we consider that Investors always have access to Trustees' reputation.

As far as the evolution of the population goes, in the study by Manapat et al. [3], an evolutionary process is used, which according to the authors can be interpreted as genetic evolution [31]. This evolutionary process causes higher payoff strategies to widespread and lower payoff strategies to eventually disappear.

Since the evolution of the population occurs according to the evolutionary process mentioned above, the highest payoff strategy tends to dominate. For the trustees this highest payoff corresponds to a value of r as close as possible to $1/b$ so that the investors still make the transfer and, simultaneously, the trustees keep the maximum amount of money possible, $r = 1/b + \epsilon$ is a Nash equilibrium [3].

Results from [3] for the different levels of trust, return, and payoffs (for both investors and trustees) as a function of the information q, can be seen in Fig 3.3. The first thing we should note is that when investors have no information, $q = 0$, evolution leads to the classical Trust Game Nash equilibrium mentioned above, *i.e.* the investors never make the transfer and the trustees never return anything,

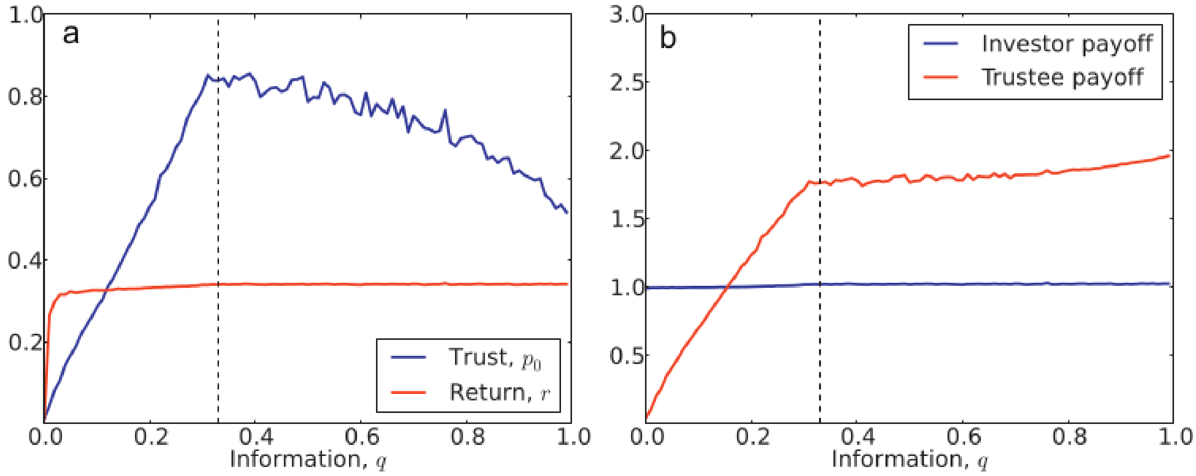

Figure 3.3: (a)The average investor trust and trustee return as functions of the information. (b) Average payoffs of investors and trustees as functions of the information. From [3]

$r = 0$. By looking at (a) we can clearly see that, as q increases, so does the return; since now more investors know the trustees' r, they must increase r so that $r > 1/b$, otherwise their payoff will diminish since fewer investors will make the transfer. As a consequence of this, the trust value increases as well. Regarding the average payoffs, however, we can see that giving information to the investors benefits trustees more than investors.

3.2 Networked Trust Games

As it was previously said, one of the focus of our work is to study the influence of reputation in networked trust games. For that, we are going to use scale-free networks, which capture important characteristics of real-world networks such as heterogeneous degree distributions *i.e.* there is a big variability between degrees in the network as opposed to a (more) homogeneous degree distribution where all the degree values are closed to the average. In order to do this, and because we could not find studies that explored the influence of both structured populations and reputation, we will analyze works with networked trust games that do not consider reputation, namely [4–6].

Abbass et al. [27] proposed an evolutionary N-player trust game with an unstructured population consisting of investors, trustees who are trustworthy, and trustees who are untrustworthy. In their study, they concluded that even though the optimal solution for the population includes investors as part of it, the evolutionary dynamics converge to a population with no investors and only trustees (of both kinds). The exception to this occurs when the initial population does not have any untrustworthy players. In [4] the authors use a population consisting of the same three types in an attempt to see whether trust can be promoted when the population is structured, namely a specific spatial topology or a social network.

Figure 3.4: Example of a social network with 11 players and the payoff of some of them, with $k_I^* = number\ of\ investors$, $k_T^* = number\ of\ trustworthy\ trustees$ and $k_{UT}^* = number\ untrustworthy\ trustees$, connected to the player in question. From [4].

Because in their study [4] populations have two different types of trustees, two different multipliers were also used, *i.e.* the factor by which the investors' money is multiplied before reaching the trustees, R_T for the trustworthy trustees and R_U for the untrustworthy, with $1 < R_T < R_U < 2R_T$. A temptation to defect investors' trust ratio r_{UT} is also introduced, with $r_{UT} = \frac{R_U - R_T}{R_T}$, which we will use in order to make it easier to analyze the results obtained.

Regarding the network topologies, Chica et al. [4] consider a regular lattice, scale-free (SF) networks with different densities as well as Erdös-Rényi (ER) random networks. For the regular lattice, the authors opted for a standard regular lattice with $\langle k \rangle = 4$ (Von Neumann neighbourhood) and periodic boundary conditions, which means all players have exactly the same degree. The SF networks were chosen as the default network for the study since these have been widely used in studies of evolutionary games. The synthetic SF networks were generated with the Barabási–Albert algorithm [24] which allows the generation of networks with different densities.

The algorithm starts with a small number m_0 of vertices, every time step a new vertex is added with *m* edges, $m < m_0$, linked to *m* different vertices that already belong to the graph. These vertices are chosen with a probability proportional to their degree, this way assuring the preferential attachment component. After *t* time steps the model results in a random network with $t + m_0$ vertices and mt edges. To create networks with different densities and $\langle k \rangle$ in [4], the authors used different values of $m \in \{2, 3, 4, 6, 8\}$. Most of the times, the authors use the SF network with $m = 3$ since this generates a network with $\langle k \rangle = 4$, which is the same as the regular lattice and the ER network. In Fig. 3.4 an example of 11 players in a social network and the corresponding payoffs of some players can be seen.

The ER network is generated by the ER model [32], the authors use a constructive random generator algorithm that creates a network with the required $\langle k \rangle$ value, in this case, 4 so it is easier to compare with the other topologies, and a binomial degree distribution. The generator, at each time step, adds a

Social Network	Density	$\langle k \rangle$	Clustering coefficient
SF2 ($m = 2$)	0.003	2.994	0.015
SF3 ($m = 3$)	0.004	3.978	0.021
SF4 ($m = 4$)	0.005	5.019	0.031
SF6 ($m = 6$)	0.007	6.986	0.040
SF8 ($m = 8$)	0.009	9.056	0.060
REGULAR	0.004	4	0
ER4	0.004	4.021	0.005

Figure 3.5: Details of the Social Networks from [4].

new node to the network and connects it to each of the other nodes with probability p. This results in a random network with very few nodes with a high degree and very few nodes with a low degree.

In Fig. 3.5 some details of the social networks mentioned can be seen.

Additionally, in order to be able to compare their results with [27] in a more meaningful way, the authors also consider a version with an unstructured population, which corresponds to a fully-connected network.

Through an evolutionary update process, players can change their strategy every time step of the game. The authors use one of the most common update rules, the proportional imitation rule, since it brings, for large well-mixed populations, the evolution of replicator dynamics, therefore, making it suitable to compare results with [27].

The proportional imitation rule is a pairwise and stochastic update rule. At any time step *t* a player may adopt one of the strategies from another player in its neighborhood. For instance, let us consider a player *a*, at time step *t* one of *a*'s neighbours is selected at random. Let us say, for example, that player *b* was chosen. First, the rule will evaluate if *b*'s net wealth at $t-1$ was higher than *a*'s. If it was not, *a* will keep its strategy, otherwise it will adopt *b*'s strategy with a probability $prob_a^t b$ dependent on both players' net wealth w at time step $t-1$.

$$prob_a^t b = \frac{max\{0, w_b^{t-1} - w_a^{t-1}\}}{\phi} \tag{3.1}$$

Where ϕ corresponds to the difference of maximum and minimum possible individual net wealth between any two players at time step $t-1$.

Concerning the experiments and results in [4], the authors use a population size of 1024 and run the model simulations for 5000 time steps each. Several values for the temptation to defect ratio r_{UT}, as well as the different network structures mentioned above and various initial population conditions, were considered in order to study the evolution of trust and global net wealth in different setups. For the r_{UT}, since its value regulates the game difficulty, three different values were mainly used, creating three different versions of the game: the easier version with $r_{UT} = 0.11$, medium with $r_{UT} = 0.33$ and harder with $r_{UT} = 0.66$.

For the easier version of the game, we can observe that trust can be promoted in all the different networks tested, investors and trustworthy trustees only disappear when the initial population is clearly dominated by untrustworthy trustees. Consequently, this results in, not only high levels of trust, but also high levels of global net wealth. These values are particularly high for the regular lattice, followed by SF networks with high densities.

When $r_{UT} = 0.33$, *i.e.* medium difficulty, the initial population, in order to promote trust, needs to have a higher number of investors and trustworthy trustees. It is interesting to note that, conversely to the previous case, SF networks are better for promoting trust, specifically SF with lower densities.

Lastly, for the harder ($r_{UT} = 0.66$) version of the game, trust can only be promoted when there are no untrustworthy trustees in the initial population.

We will now continue exploring trust games, but this time focusing on the effects update rules have on networked N-player trust games, mainly by analyzing the work done by Chica et al. [5] which in terms of types of agents and payoff rules, uses the same N-player networked trust game as in [4].

As explained before, players can change their strategy every time step of the game through an evolutionary process based on neighbour imitation. In [4] this imitation occurred through a rule called Proportional imitation (PROP). In [5], however, three more update rules are considered. In order to better understand the influence of considering different update rules, we will show what these consist in.

First, the authors present the most simple update rule, named Unconditional Imitation (UI) [21]. With that rule, a player a, every time step, copies the strategy of its highest payoff neighbour b if b's payoff at time step $t - 1$ is better than a's at the same time step. This rule is, therefore, totally deterministic.

Secondly, the authors discuss the Hybridization of UI and a voter model [33] (UI-VM). The voter model is a stochastic process, every time step a player a imitates the strategy of one of its neighbours at random. In order to have this hybridization of UI with the voter model, a parameter q is introduced. A player a will choose the voter model with a probability q and UI with a probability $1 - q$. In [5] the authors set $q = 0.1$.

Lastly, the Moran process (MO) [34] is introduced. With MO, at each time step t a player a evaluates all its neighbours' payoffs at time step $t - 1$ and assigns them probabilities proportional to their payoffs values. After this player a will then imitate one of the strategies of its neighbours according to the probabilities previously assigned.

Regarding the experiments and results in [5], the authors used very similar initial conditions to [4], a population size of 1024 agents, with simulations of 5000 time steps. Once again r_{UT} was used to regulate the difficulty of the game with $r_{UT} \in \{0.11, 0.33, 0.66\}$. As far as network topologies, this time only a regular lattice and a SF network, created using the Barabasi-Albert algorithm with $\langle k \rangle = 4$, were considered.

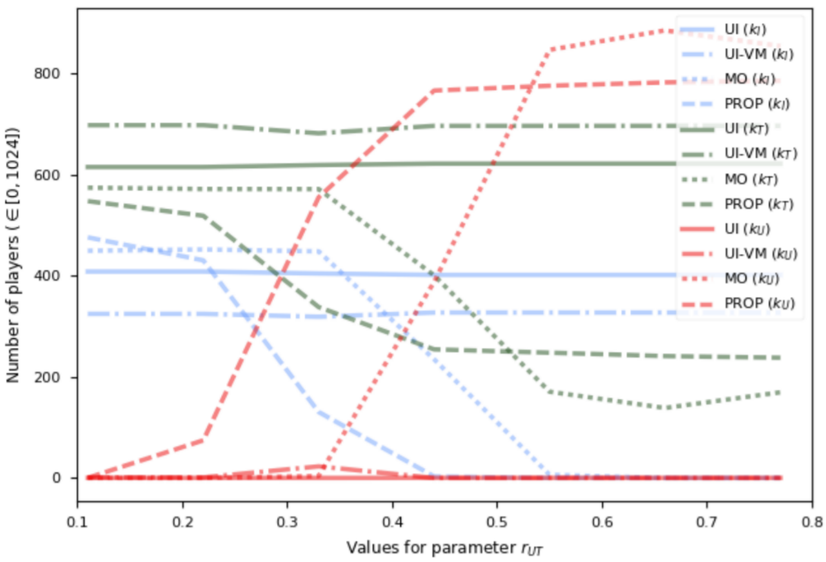

Figure 3.6: Number of players of each strategy (k_I, k_T, and k_U) at the end of the simulations, when using a regular 32x32 lattice From [5].

Using the global net wealth of the population as the evaluation metric, analyzing the results, we can see that when the game is easier ($r_{UT} = 0.11$), for the SF network, MO and PROP get the best results. For the regular lattice, all the update rules get very high results.

When $r_{UT} = 0.33$, for the SF networks, only UI and UI-VM perform well in terms of promoting trust, however, for the regular lattice, only PROP prevents trust from evolving.

Lastly, when the game becomes harder ($r_{UT} = 0.66$), MO and PROP cannot promote trust for any of the network topologies. UI and UI-VM are able to get high global net wealth values for different initial population conditions if considering the regular lattice.

In Fig. 3.6 we can see the number of investors, trustworthy trustees, and untrustworthy trustees at the end of the simulations, for the different update rules, in a regular lattice. In summary, this works shows that there is a non-trivial relation between update rules and trust.

We will now explore the final work in this chapter which consists of an attempt to model trust, using evolutionary game theory, in a sharing economy [6]. This study is particularly interesting because it considers the same four possible strategies as in our own model. Reputation is not considered, however, which we believe is a limitation as many of the real world cases of the sharing economy have some kind of reputation system implemented *e.g.* Airbnb or Uber.

In the trust model developed by Chica et al. [6] players have four possible strategies: being a trustworthy consumer (TC), being an untrustworthy consumer (UC), being a trustworthy provider (TP), and being an untrustworthy provider (UP).

Concerning the social network used, for this study, the authors only considered a real network that

Providers	Consumers TC	UC
TP	R, R	$-S$, $Temp$
UP	X, $-X$	X, $-X$

Figure 3.7: Payoffs for the sharing economy trust game. From [6].

corresponds to the email network from a university in Tarragona, Spain.

The net wealth of individual players is calculated using their payoffs which vary accordingly to the strategies adopted by their neighbours. The interactions are pairwise, and the total net wealth of each player is calculated by adding all the payoff values from each interaction with its neighbours. If either two consumers or two providers interact there is no resultant payoff. In Fig. 3.7 we can see the payoff for all the interactions, where $2R > Temp > R > S > X$.

The players' strategies can change every time step of the game through an evolutionary process. For this work, the authors opted for the proportional imitation rule as the only update rule.

Regarding the experiments and results, the simulations were run with a population of 1133 players, for 5000 time steps each run. The payoffs were calculated using the matrix in Fig. 3.7 with $Temp = 40$, $S = 20$ and $X = 10$. The parameter R was used to control the difficulty of the game and was set with values ranging from 21 to 39.

In order to see the impact of changing the initial conditions in the final results, 5 different scenarios, varying the initial proportions of each strategy in the population, were explored.

For $33 \leq R \leq 39$, which corresponds to the easier version of the game, the population is dominated by TC and TP players for every scenario with the exception of when the initial number of TP and TC was really small (5% of the population each).

Changing the game to a moderate difficulty, $25 < R < 33$, the final number of trustworthy players, consumers and providers, decreases and consequently the final number of untrustworthy players increases.

Lastly, when considering a harder difficulty, $21 \leq R \leq 25$, the results were rather unexpected. One would expect, particularly when considering the results for the moderate difficulty, that the untrustworthy players would continue to dominate, however, this is not the case. First, for all the scenarios with the exception of the one where the initial number of trustworthy players was very low (5% of the population of TC and 5% of TP), the number of UC players actually decreased. Second, for both the scenarios where the initial number of UC players was very small (5%), the number of TC and TP players increased.

4

Model

Contents

As mentioned in chapter 3, evolutionary game theory has already been used for modeling trust, using slightly altered versions of the Trust Game, both in unstructured populations, *e.g.* [3, 27], and with networks, *e.g.* [4–6].

Sigmund in [2] studies the replicator dynamics of the reduced Trust Game using a similar version of the Trust Game introduced by Berg et al. [1], *i.e.* considering a very similar payoff matrix. For this version the results obtained correspond to the game's Nash Equilibrium, *i.e.*, in the long run, the population will consist only of players who always defect. Later, in the same work, another version considering reputation effects is introduced yielding different results, favoring the promotion of both trust and trustworthiness.

The version of the Trust Game making use of reputation in [2] considers infinite unstructured populations. As mentioned in chapter 1, many times, from a mathematical point of view considering infinite populations is more convenient, however, real world populations (namely those where individuals play the Trust Game) are finite and individuals interact over social networks. Considering finite populations can introduce considerable changes [10]; one that we should note is the fact that, in finite populations, strategies that form Nash Equilibria, or that are evolutionary stable, may not always be highly prevalent [10, 35].

Based on the work of Kandori [9] and Nowak and Sigmund [36], Manapat et al. [3] used a different reputation system with unstructured finite populations where, for some interactions, the investor learns information about the trustee before the interaction occurs and takes this information in consideration when making the decision whether to make the transfer or not.

Our model aims to implement an altered version of the Trust Game where we consider the effects of reputation in finite populations as in [3]. However, the reputation aspect will be added in a different way, similar to the one used in [2], *i.e.* while in [3] investors only have information on the trustees with a certain probability, in our model and in [2] Investors always have access to Trustees' reputation. Furthermore, instead of only using unstructured populations, we will also consider a networked version of the Trust Game similar to the one in [6].

We use this model to study the evolution of trust making use of computer simulations and evolutionary game theory, in the contexts previously described, by creating an evolutionary process where the strategies of the investors and trustees change through time. This change is done via a mechanism of social learning [37], where players can imitate the strategies of those performing better, causing higher payoff strategies to proliferate while lower payoff strategies diminish.

4.1 Game Payoffs

The Trust Game is played by all the players for a finite number of time steps. At each time step of the simulation, players interact with each other accordingly to the type of population structure considered. These interactions between the players result in a certain payoff for each player. Every time step of the simulation new payoffs are calculated for every player since players might have changed their strategy in the previous time step, hence resulting in new payoffs in the present one.

A player's (total) payoff for any time step of the simulation corresponds to the sum of the payoffs from every interaction that player was part of in that time step. The payoffs of any interaction are calculated according to the payoff matrix in Table 4.1, which corresponds to Table 5.28 in [2] (when the parameter μ in the book is 0). We should also note that Table 4.1 corresponds to Table 2.2 with some added complexity: the cost for a Trustee to cooperate is different from the reward it has to a cooperative Investor; Defective Investors have a payoff of 0 when playing against defective Trustees; Cooperative Investors have a negative payoff when playing against defective Trustees.

Table 4.1: Trust Game Payoff Matrix. Corresponds to Table 5.28 in [2], when $\mu = 0$

	$\mathbf{f}_1$	$\mathbf{f}_2$
$\mathbf{e}_1$	$(\beta - c, b - \gamma)$	$(-c, b)$
$\mathbf{e}_2$	$((\beta - c)\upsilon, (b - \gamma)\upsilon)$	$(0, 0)$

This payoff matrix can be understood as following: players can act as one of two roles in each interaction: either as an investor (row) or as a trustee (column). The first position of each cell corresponds to the investor's payoff, while the second position payoff corresponds to the trustee's. An investor may choose to make a transfer, *i.e.* cooperate ($\mathbf{e}_1$); or to defect ($\mathbf{e}_2$). The same applies to trustees who can either return a certain amount to the investor, *i.e.* cooperate ($\mathbf{f}_1$); or defect ($\mathbf{f}_2$), *i.e.* do not transfer back anything. In this version of the trust game, if an investor decides to cooperate, then he will donate a sum c to the trustee, which will be multiplied by a factor $r > 1$ resulting in $b > c$. The trustee, in turn, will return an amount β to the investor, costing him γ, if he decided to cooperate, and 0 otherwise (no cost in this case). We assume $0 < c < \beta$ and $0 < \gamma < b$. Lastly, the variable υ corresponds to the likelihood that defective investors cooperate if they know that they will be rewarded. This means essentially that υ corresponds to the trustees' reputation, *i.e.* the probability that a player with a strategy $\mathbf{f}_1$ becomes known as a cooperator, and consequently a player with a strategy $\mathbf{e}_2$ cooperates too. Thus, defective investors ($\mathbf{e}_2$) can either have payoff of $(\beta - c)\upsilon$ when interacting with a trustee with a strategy $\mathbf{f}_1$ or 0 if interacting with a trustee with a strategy $\mathbf{f}_2$, *i.e.* no costs or benefits.

4.2 Evolutionary Update

When playing several instances of Trust Games in a population, the players' strategies can change through an evolutionary process that can be interpreted as social learning as mentioned above. It is named social learning due to the fact that players can imitate the strategies of others in the same population. In fact, people often resort to such a strategy update process [38].

In our model, each round, firstly one player a is randomly selected, then mutation occurs with a probability μ (parameter of the process), *i.e.* there is a small chance the chosen player will just change his strategy to a random one (might be the same he already had). If mutation did not happen, then a new player b from the same population is randomly selected (investors can only imitate other investors and trustees can only imitate other trustees) accordingly to the population structure considered. A pairwise comparison rule was then adopted in order to calculate the probability (p) of the first player (a) imitating the second player (b) based on both of their resultant payoffs in that time step. In our work, as in [3], we use the Fermi function as this pairwise comparison rule, as studied by Traulsen et al. [39]:

$$p = \frac{1}{1 + e^{-\beta(\pi_b - \pi_a)}} \tag{4.1}$$

The variables π_a and π_b correspond to the accumulated payoff of player a and player b respectively, calculated for each player as the sum of all his interactions' payoffs. Let us consider here the parameter β as the intensity of selection (not to be confused with β introduced in the context of the Trust Game payoff, as the Trustee return). This means imitation of strategies will occur with a probability proportional to the difference between both players' payoff (for $\beta > 0$), and that if β increases so does the dependence on this difference.

For the symmetric versions of the game, *i.e.* where players play as both roles and therefore have two independent strategies, players will only imitate one of the other player's strategy.

4.3 Unstructured Populations

Sigmund in [2], studies the replicator dynamics of two versions of the Trust Game: a reduced version similar to the one introduced by Berg et al. [1]; a second version where reputation is introduced. Since the scope of this work is the effects of reputation in the Trust Game, our main focus will be on this second version of the game.

The results in [2] are for infinite populations, however, in the real world populations are finite. Our objective is to, using the same payoff matrix (Table 4.1), simulate the same interactions in a finite population context.

We first consider the case where there are two populations, one of each role (investors and trustees); at every time step of the simulation, each individual from either of the populations interacts with all of the individuals from the population with a different role. In this case, the imitation process occurs only between individuals of the same population, and every player can imitate any of the others.

Secondly, we consider a symmetric version of the game, where there is only one population; all the individuals play as both roles and may have different strategies for each one, each individual has two independent strategies. Players interact with the entire population, playing in both roles, and the total payoff of each player corresponds to the sum of his total payoff as an investor with his total payoff as a trustee. Once again, every player can imitate any of the others, however, they can only imitate strategies of the same role, *i.e.* if a player is imitating another player's strategy, for instance as an investor, he may only change his strategy as an investor as well.

4.4 Structured Populations

Many of the most important studies regarding the Trust Game, as in the first part of this Thesis, consider well-mixed (unstructured) populations. This assumption is far from a real world situation where populations have a finite number of individuals and a structure.

For this part of our study, players are placed in a social network, which will have an impact on the way information is spread, and how the pairs to play games are formed. Regarding the network structure, we use scale-free networks since they capture important characteristics of real world social networks, such as a highly heterogeneous degree distribution. In fact these networks have been used extensively in studies related to evolutionary games (*e.g.* [40]), namely in the case of networked Trust Games like [4] and [5].

By using networks, interactions between players are constrained by the network topology, *i.e.* every player will only interact with his direct neighbours. For most of our experiences, players will play in both roles, one at a time, every time step; their total payoff will correspond to the sum of the payoffs when playing in both roles. The imitation process, analogous to the interactions, will be restrained to a player's direct neighbours. Like in the unstructured symmetric version of the game, each player will have two independent strategies (one as an investor and one as a trustee) and can only imitate strategies of the same role.

4.4.1 Asymmetric role assignment

As previously mentioned, our default setting for the networked version of the Trust Game is for all the players to play as both an Investor and a Trustee. The idea behind asymmetric role assignment version of the game, however, is to, for every interaction, force a dependency between the role a player has in that interaction and his characteristics, namely his degree in the network.

For each interaction, the role assignment for any player may be different, depending on the degrees of the players taking part in that interaction, therefore each player will still keep two independent strategies (one for each role).

Every time any player a interacts with any player b, a will act as an Investor with a probability p_i, calculated accordingly with the following equation:

$$p_i = \frac{k_a^\lambda}{k_a^\lambda + k_b^\lambda} \tag{4.2}$$

Here, k_a and k_b correspond to player a and player b's degree in the network respectively. The variable λ, which may take negative values, controls the dependency between the degree of a player and his role. If λ is 0 then the player's role is uniformly random, *i.e.* the role of investor is attributed to one of the agents with the same probability. The higher λ is, the more likely the player with the larger degree is to act as an Investor. For negative λ values, the lower λ is, the more likely the player with the higher degree is to act as a trustee. With this experience, our main interest is to see the consequences of forcing most of the individuals that are hubs in the network to have a certain role - which happens with either considerably large or small values for λ - regarding the promotion of Trust and Trustworthiness.

4.4.2 Diversity in the reputation

In the previous versions of the networked Trust Game considered, all the players have the same reputation, *i.e.* the same v value in the payoff matrix (Table 4.1). In this version, our aim is to determine the effects of varying the value of v, while always maintaining the average v in the population the same.

We shall then consider 3 scenarios to compare the effects of this diversity in reputation: the baseline where the whole population has the same v; a second one where we consider two different v values and assign the larger to the half of the population with a higher degree and the smaller to the remaining half; another one where we consider the same two different v values and assign each of them to half of the population, randomly selected. This allows implementing the situations in which highly connected nodes are also subject to higher scrutiny from overall population and test their impact on the evolution of trust.

4.4.3 Hybrid societies with pathological players

In order to avoid initial imbalances in terms of strategies in the population, the default setting before starting the simulation is to randomly assign strategies to each player, while making sure that 50% of the individuals have one strategy (either cooperative or defective) and the remaining the other. For this version of the Trust Game, however, we introduce pathological players, *i.e.* players that, regardless of what happens during the simulation and of the role they are assuming, will always cooperate for every interaction. The introduction of pathological players is particularly interesting since, in some real world cases, it may exist the possibility, using artificial intelligence and socially interactive agents to introduce hard-coded behaviors, such as trust and cooperation, in a population.

In order to compare the effects of the pathological players being network hubs or not, we consider two variations: Firstly, one where we define several thresholds for the degree of the nodes in the network. Players with a degree above these thresholds will be pathological players; Secondly, we count the number of pathological players assigned for each of the thresholds considered in the first scenario and assign the same number of pathological players but by selecting them randomly out of the population.

5

Results and Discussion

Contents

As mentioned in chapter 4, the experiments done consist of applying our model to computer simulations. To do this we wrote and simulated a program in Python 3.6.8.

In this chapter we present and discuss the results from these computer simulations, firstly for the unstructured populations and secondly for a scale-free network (structured populations) where we divide the results considering the three scenarios previously mentioned.

5.1 Methods

In our simulations, depending on the experiment we are considering there are some small changes in its initial setup. In the non symmetric version with unstructured populations, we consider two populations of 500 individuals each. For the symmetric version with unstructured populations and all the versions with networks, we consider a single population with 500 individuals. The networks considered are scale-free networks with an average degree of 4.

Every individual is represented by his payoff, his strategy as an investor, and his strategy as a trustee (with the exception of the non-symmetric version where each individual only plays in one role and therefore only has one strategy).

Every simulation has 10^5 rounds, consisting of one run. Regarding the stream plots for the unstructured populations we consider the state of the population after a single run of the whole experience, however, transition probabilities correspond to an average of 10^3 runs for each of the possible combinations of states. All the other results correspond to an average of 200 runs for each experience.

The players' payoffs, as it was previously mentioned, are calculated according to the payoff matrix in **Table 4.1**. The default values for the payoff matrix are: initial stake of a cooperative investor $c = 1$; amount returned by a cooperative trustee $\beta = 3$; initial stake multiplied by factor $r = 3$ corresponds to $b = 3 \times 1$; cost to a cooperative trustee $\gamma = 2$; and trustees' reputation $\upsilon = 0.5$. The values for the variables c, β and γ are fixed for all the versions of the game. The remaining variables also have the same values for all the versions with the exception of: the asymmetric role assignment version, where we consider multiple values for the multiplication factor r (value by which the initial stake of an investor is multiplied before reaching a trustee); the diversity in reputation version where we consider different values for υ (the trustees' reputation). The payoff matrix with these default values can be seen in **Table 5.1**.

Table 5.1: Trust Game Payoff Matrix with default values

	$\mathbf{f}_1$	$\mathbf{f}_2$
$\mathbf{e}_1$	$(2,1)$	$(-1,3)$
$\mathbf{e}_2$	$(1,0.5)$	$(0,0)$

From **Table 5.1** we can conclude that there is no Nash equilibrium in pure strategies. It is possible, however, to compute the Mixed-Strategy Nash Equilibria. Let us consider that Investors play with a strategy $\mathbf{e}_1$ and $\mathbf{e}_2$ with probability q and $1-q$ respectively; Trustees play with a strategy $\mathbf{f}_1$ and $\mathbf{f}_2$ with probability p and $1-p$ respectively. Then we can calculate the q and p that make individuals, playing in a given role, indifferent between the two strategies

$$q + 0.5(1-q) = 3q$$

$$2p - 1(1-p) = p$$

Thus, $q = 0.2$ and $p = 0.5$ and the mixed strategies $(0.2; 0.8)$, for the investors, and $(0.5; 0.5)$, for the trustees, are a Nash equilibrium.

Lastly, concerning the evolutionary update, the probability of a random mutation occurring before imitation takes place is $\mu = 0.01$. The imitation occurs according to the Fermi function (equation 4.1), as defined in chapter 4, here, the default value for the intensity of selection is $\beta = 10$.

5.2 Results

Below we present the results regarding the computer simulations with the Trust Game. As mentioned, results are divided according to the type of populations considered: firstly we present the outcome of unstructured populations playing the Trust Game, considering two main scenarios, a non symmetric version of the game (individuals only play in one role) and a symmetric one (individuals always play in both roles); secondly, we consider structured populations, namely in a scale-free network, where we look at the three different scenarios described in chapter 4: 1) asymmetric role assignment; 2) diversity in reputation, and 3) hybrid societies with pathological players.

As mentioned in chapter 4, we consider the payoff matrix presented in [2], which was only studied regarding infinite well-mixed populations. Although the main focus of this work is to study the effects of individuals playing a slightly altered version of the Trust Game, namely by introducing reputation, for the versions with unstructured populations, we first start by trying to prove that the results in [2] are extendable to finite (unstructured) populations by comparing the evolutionary dynamics of both. Furthermore, we calculate the average values for both trust and trustworthiness, *i.e.* out of the 500 individuals in the population how many are cooperators when they play as an investor and how many are cooperators when they play as a trustee respectively, in order to compare the promotion of trust and trustworthiness between unstructured and structured populations.

5.2.1 Unstructured Populations

Firstly we considered the reduced Trust game in [2] which is very similar to the original version of the Trust Game introduced in [1] (does not take reputation into account). The results for this version show that, after the initial state, all the players become defectors, regardless of the role they are assuming.

Using our model we calculate the evolutionary dynamics for both the non-symmetric and symmetric versions of the reduced Trust Game, by simulating the game in all the possible states of the population, *i.e.* all the possible combinations of the players' strategies, and estimating for each one the most likely state it will transit to.

The figures resultant from these simulations (stream plots) represent the most likely direction of evolution, Fig. 5.1 shows the evolutionary dynamics for the non symmetric version of the reduced Trust Game and Fig. 5.2 for the symmetric one. Once again, and confirming our previous results, both figures show that, regardless of the initial proportions of players with defective and cooperative strategies, the population tends to evolve to a state where all the players become defectors. For instance, looking at Fig. 5.1 if we consider an initial population with 500 defective Investors and 500 cooperative Trustees (bottom right corner of the figure) it will most likely initially gravitate to a state with fewer cooperative trustees and about the same number of cooperative and defective investors. However, if we continue playing the game, then the most likely final state of the populations is for all the players to be defective (top right corner of the figure).

Next, we consider the Trust Game version that takes into account reputation. The results, for both the non-symmetric version in Fig. 5.3 and the symmetric version in Fig. 5.4 were similar to the ones in [2] (Fig. 3.2), showing that they are indeed extendable to finite populations and that reputation does have a positive effect regarding the promotion of trust and trustworthiness.

We can better observe this positive effect by looking at both stream plots for the Trust Game with reputation (Fig. 5.3 and Fig. 5.4) and comparing them with the version with no reputation (Fig. 5.1 and Fig. 5.2). Although most of the transitions on the left side of the figures are similar, when we get closer to the state with no cooperative individuals (top right corner) we can see that for the Trust Game version with reputation, the most likely direction of evolution is not to stabilize there, but instead to evolve to a state with more cooperative trustees, thus forming a cycle. We can better understand this cycle by looking at the payoff matrices for both versions of the game: as mentioned before, for the version with no reputation, the Nash Equilibrium is for both Investors and Trustees to have a defective strategy, justifying the stabilization of the population in this state (top right corner of both figures); for the version with reputation, however, there is no Nash equilibrium in pure strategies and when the population reaches the state where most of the Investors and Trustees have a defective strategy then it becomes better for the Trustees to cooperate again, since the introduction of reputation causes them to have a higher payoff when cooperating rather than defecting against defective Investors.

In order to verify the effects of introducing reputation into the model, we calculate the average number of individuals with cooperative strategies, over time, which allows to quantify the level of trust and trustworthiness in the population in these simulations. For the non symmetric version the average number of cooperative Investors and Trustees was 108.38 and 323.84 respectively. For the symmetric the results are rather similar, with an average 106.74 of cooperative investors and 322.49 cooperative trustees.

As mentioned, these results match the ones obtained studying, analytically, a replicator equation. The fact that here we resort to simulations and, numerically, calculate the most likely direction(s) of evolution, allows us to extend the approach to structured populations, as we shall do in the next section.

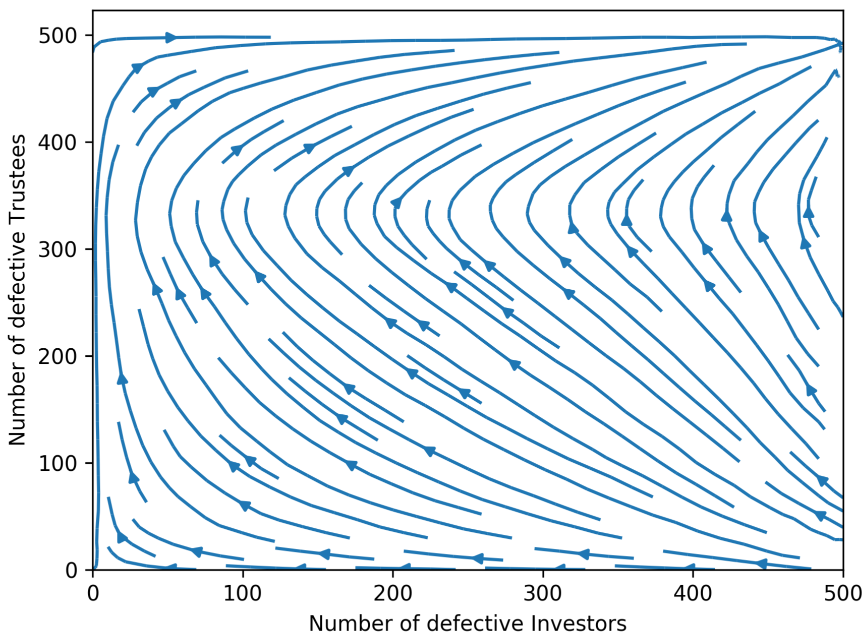

Figure 5.1: Evolutionary Dynamics of the Non Symmetric version of the Trust Game with no reputation for Unstructured Populations

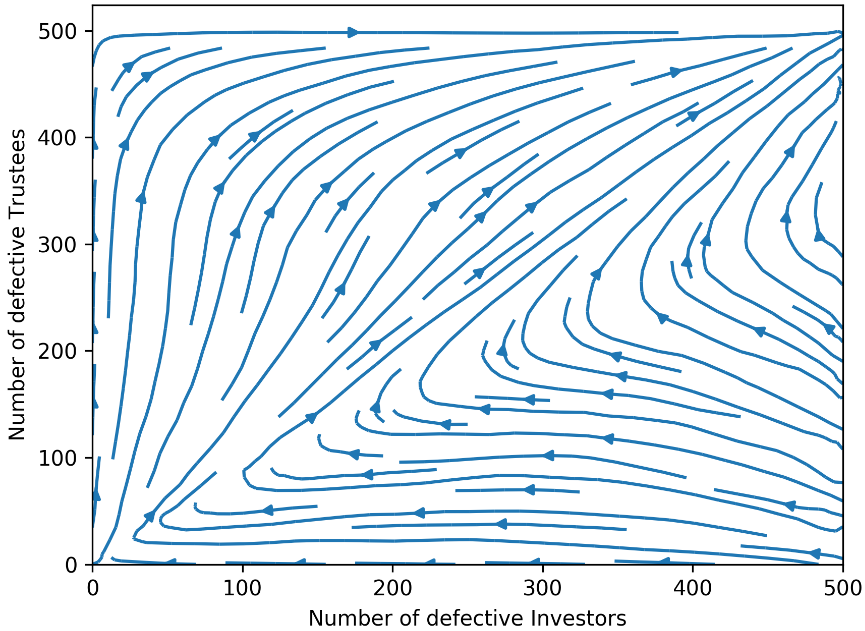

Figure 5.2: Evolutionary Dynamics of the Symmetric version of the Trust Game with no reputation for Unstructured Populations

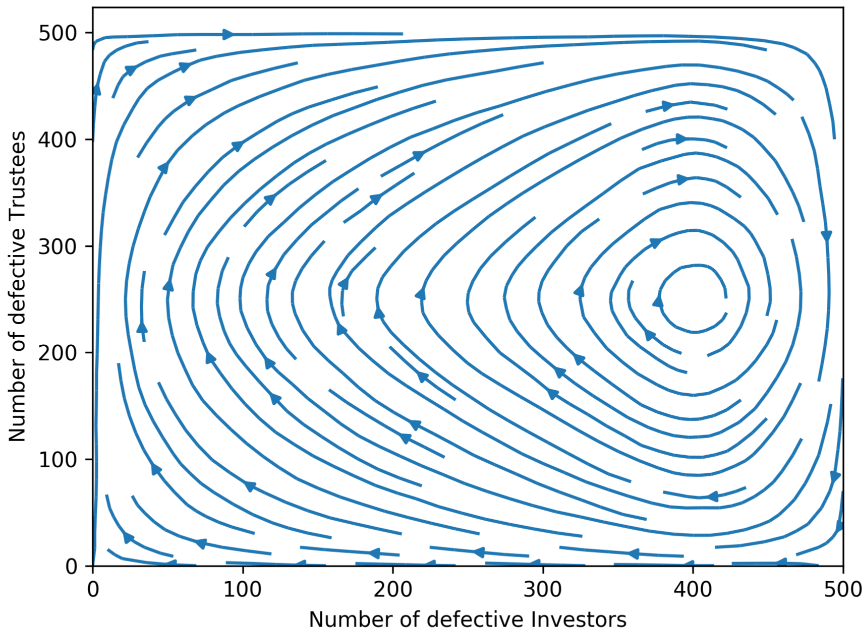

Figure 5.3: Evolutionary Dynamics of the Non Symmetric version of the Trust Game with Reputation for Unstructured Populations

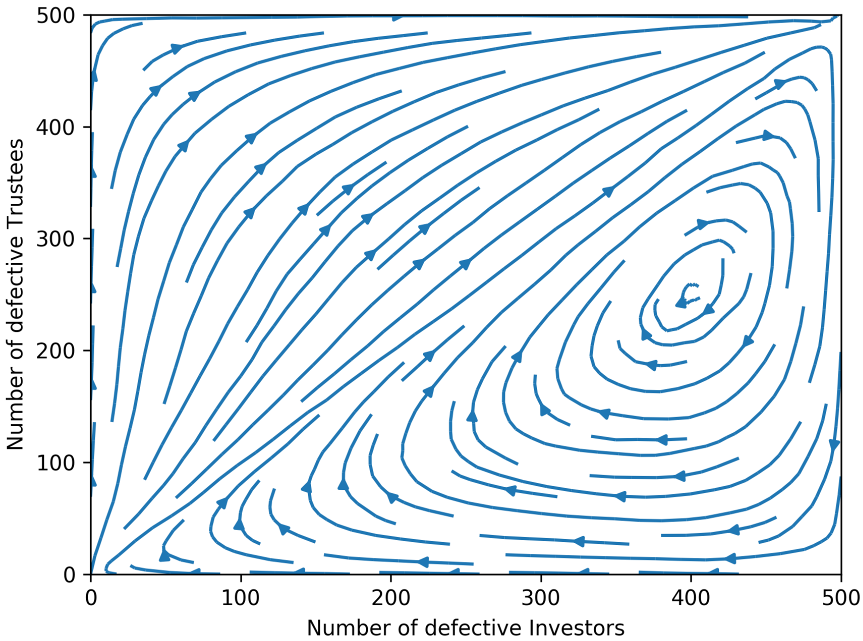

Figure 5.4: Evolutionary Dynamics of the Symmetric version of the Trust Game with Reputation for Unstructured Populations

5.2.2 Structured Populations

Although results for Trust Game with reputations and unstructured populations, revealed a clear improvement regarding the promotion of trust and trustworthiness, the average value of trust is still low when compared to behavioral data resultant from experiments with real people like in [1]. In an attempt to more accurately emulate human behaviour in real life situations we did the same simulations just mentioned, however this time, using structured populations in a scale-free network.

With the unstructured populations, we assume that all the individuals in a population interact with each other with the same probability, however, in real life scenarios individuals only interact with a limited number of people, *i.e.* populations in the real world are structured, and the emergence and promotion of cooperation may be affected in a positive way by that structure [41].

Regarding the trust game version with structured populations, as mentioned in chapter 4, we explore and divide the results into three different scenarios where we introduce some variations that we believe take place in real life circumstances.

5.2.2.A Asymmetric role assignment

So far, in the unstructured version of the game, we considered that every player always plays as a certain role (or as both roles in the symmetric version), however, in real life situations that resemble the Trust Game, not every individual acts in a certain role with the same probability. One could say that this probability of playing as an Investor or Trustee depends on an individual's degree in the network considered. For instance, if we consider the network in which nodes are Uber's drivers and Uber's clients (let us consider Uber's drivers as trustees and the passengers as investors), where a driver is linked to every client he ever had (and vice versa), we may assume that the drivers are hubs in this network, *i.e.* the individuals with larger degrees in this network represent almost always trustees.

In this version, we try to apply the idea we just described to the trust game, Fig. 5.5 shows the results regarding the trust values and Fig. 5.6 the trustworthiness values. As detailed in chapter 4, for every interaction between two players we force a dependence between the players' role in that interaction and their degree in the network. This dependence is controlled by equation 4.2, namely through the value that the variable λ assumes. In our experiments we consider the values -2, -1, 0, 1 and 2 for λ. For $\lambda = 0$ the role assignment is attributed uniformly, therefore the game is played as up until now. Higher values (1 and 2) make the player with the higher degree more likely to be an investor and lower values(-2 and -1) more likely to be a trustee.

Additionally to the role assignment variation we just described we also consider three different levels of difficulty in the game, by varying the multiplication factor of the investors' initial stake. For higher multiplication factor values, the difference, regarding the payoff, between cooperating and defecting is lower for the Trustees. As a consequence of this the Trustees will be more likely to cooperate, thus making the game easier.

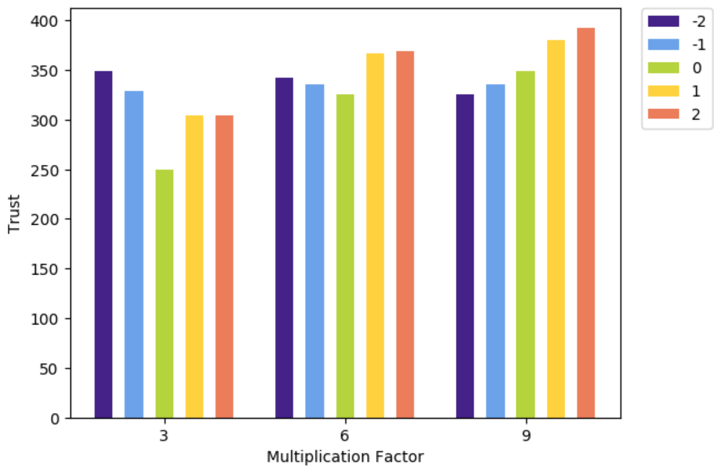

Figure 5.5: Average number of cooperative Investors over time by varying the multiplication factor (r) and λ (equation 4.2), each colour represents a different λ value

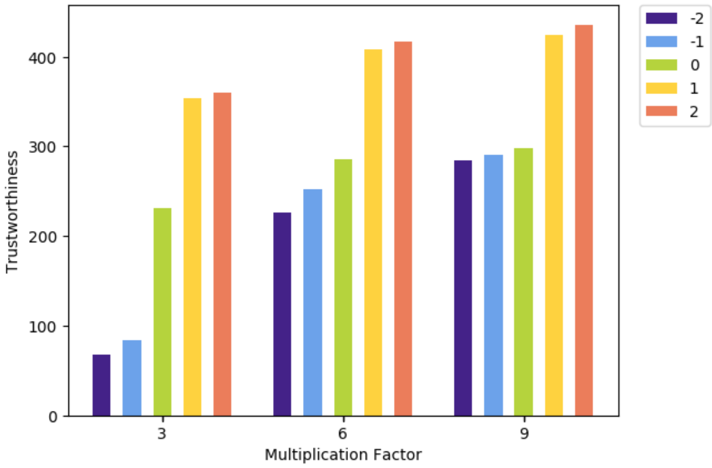

Figure 5.6: Average number of cooperative Trustees over time by varying the multiplication factor (r) and λ (equation 4.2), each colour represents a different λ value

5.2.2.B Diversity in the reputation

In the previous version, we studied the influence of an individual's degree in the network on his most probable role. In regard to reputation, however, we considered that all the players have the same value, *i.e.* if we consider Table 4.1 all the individuals have the same υ. Yet, when picturing real life situations it is easy to think that individuals with a higher degree will have, most of the times, a higher probability of having a known reputation as well. For instance, if we consider the sellers in Alibaba retail store as nodes in a network, that are linked to every client they ever had, most likely, the ones with a higher degree in the network will also have more reviews and, as such, a higher probability of having a public reputation.

In this version, we return to a symmetric game (players play as both roles and their payoff corresponds to a sum of the payoff as each role) and we consider three different situations regarding the reputation values distribution:

- a uniform distribution where all the players are assigned the exact same reputation value, $\upsilon = 0.5$
- a distribution taking into consideration players' degree in order to replicate what was just described, where we pick two different values for reputation, $\upsilon_1 = 0.1$ and $\upsilon_2 = 0.9$, assign υ_1 to the half of the individuals with the lowest degrees and υ_2 to the half with highest degrees, assuring this way that the average reputation value remains the same
- a random distribution of the same υ_1 and υ_2 values, while assuring again that the average reputation value stays the same, *i.e.* a random half of the population gets assigned υ_1 and the remaining individuals υ_2

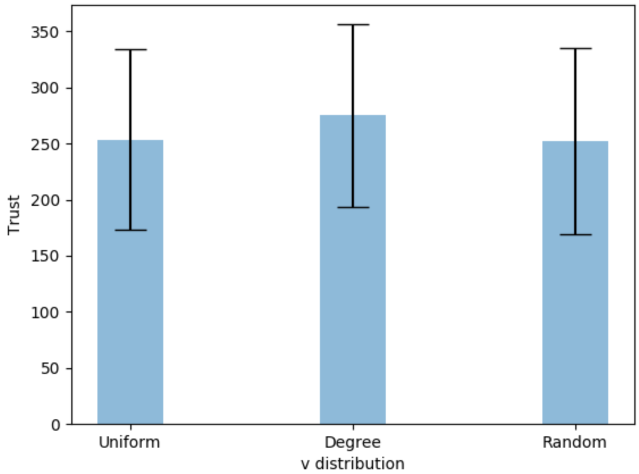

Figure 5.7: Average Trust levels and standard deviation of average values of Trust over runs by varying the way reputation is assigned

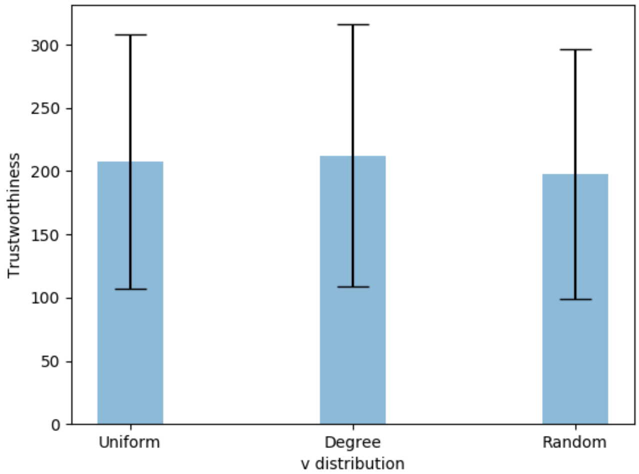

Figure 5.8: Average Trustworthiness levels and standard deviation of average values of Trustworthiness over runs by varying the way reputation is assigned

Fig. 5.7 and Fig. 5.8 show the results, for these three variations of reputation distribution, regarding the average and standard deviation values of trust and trustworthiness respectively.

5.2.2.C Hybrid societies with pathological players

In this version we follow the main idea of the previous one, however, rather than assuming that players with a higher degree have a higher reputation as well, we will instead assume that they will always cooperate, regardless of the role they are playing as, and therefore are denominated pathological players [13] (also named resilient [12]). In a real life context, this means we presume that individuals with really large degrees in the network, *e.g.* the British online retailer Asos, most likely always cooperates (in this example it would mean, never deceive the clients and always send the product they asked for) when compared to individual sellers in smaller retail shops.

Furthermore, rather than halving the population, as in the diversity in reputation scenario, we really want to focus on the network hubs. To do this in our experiment, as explained in chapter 4, we defined 4 different thresholds for players' degrees, each considered in separate runs, above which players act cooperatively for both roles. Additionally for each of the thresholds, we also consider the same number of players but randomly selected for comparison reasons, as explained in chapter 4.

Fig. 5.9 shows the results for both the pathological players chosen according to their degrees and the ones chosen randomly regarding the average trust values, and Fig. 5.10 regarding the average trustworthiness values. The thresholds for players' degrees are 10, 20, 30, and 40 which correspond to 29, 11, 3, and 2 pathological players assigned respectively. As we can see in Fig. 5.9 having only 2 hubs as pathological players already has a tremendous effect regarding the promotion of Trust in the population.

Lastly, we should note that, for comparison reasons, for the regular version with no pathological players - left column of Fig. 5.7 and Fig. 5.8 - the average trust was 254 and the average trustworthiness 208.

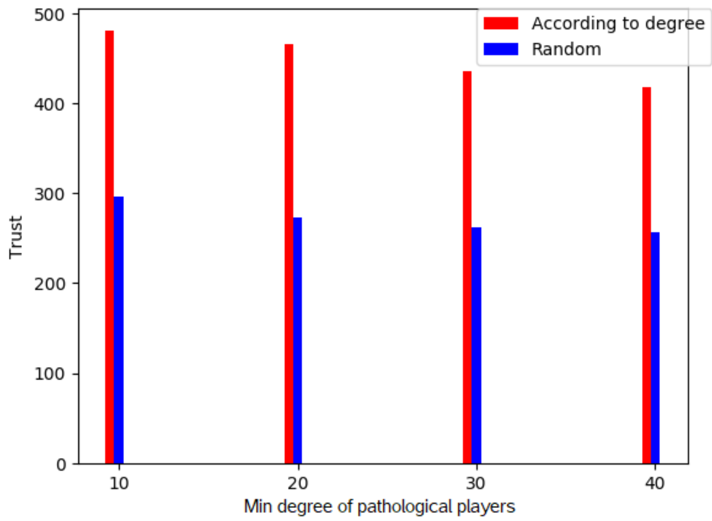

Figure 5.9: Trust levels for pathological players either selected according to thresholds for degree or randomly

Since the results for this version of the Trust Game were the best ones regarding trust and trustworthiness (if we are only considering the regular multiplication factor of 3), particularly when looking at the smaller degree threshold considered, and therefore the most pathological players, we wanted to see how this would show in the stream plots for this version of the game, *i.e.* how it affects the most likely

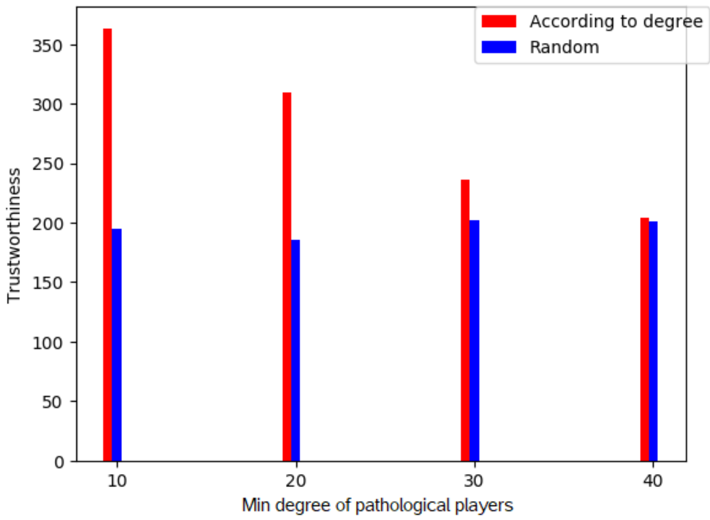

Figure 5.10: Trustworthiness levels for pathological players either selected according to thresholds for degree or randomly

direction of evolution of the population.

In order to more precisely analyze the effects of the pathological players we also consider the stream plot for the regular version of the Trust Game with reputation played in a network, which can be seen in Fig. 5.11. Fig 5.12 shows the evolutionary dynamics for the Trust Game with pathological players selected according to their degrees, considering the lowest degree threshold (29 pathological players).

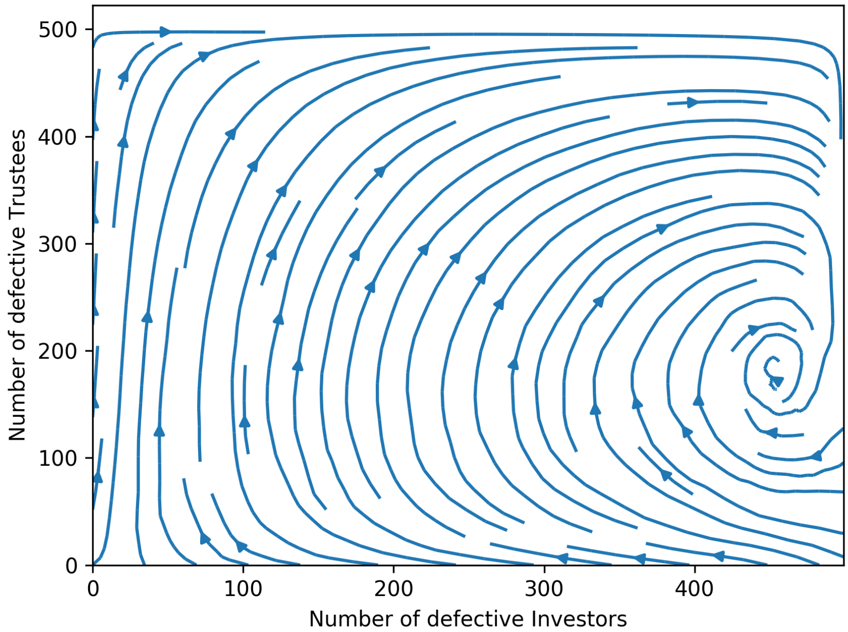

Figure 5.11: Evolutionary Dynamics of the regular version of Trust Game with reputation for Structured Populations

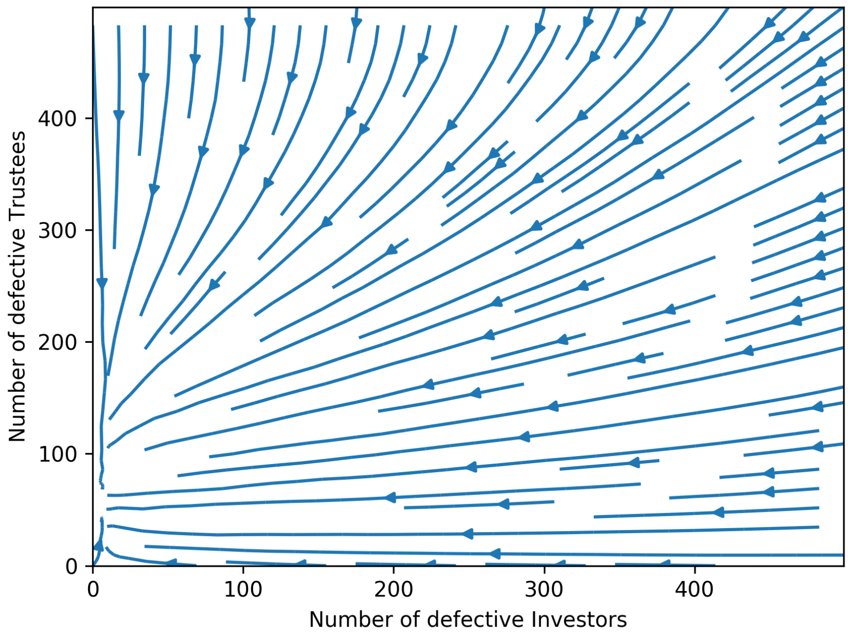

Figure 5.12: Evolutionary Dynamics of the Trust Game version with (29) pathological players selected according to their degree in the Network

5.3 Discussion

5.3.1 Unstructured Populations

Firstly, regarding the reduced Trust Game version, as predicted by the replicator dynamics and the Nash Equilibrium of the game presented in [2], for both the non-symmetric and the symmetric version of the game the results for trust and trustworthiness are 0 after the initial state of the game, *i.e.* all the players become defectors, regardless of the role they are assuming. The evolutionary dynamics results (Fig. 5.1 and Fig. 5.2) are also considerably similar to the ones in [2] allowing us to conclude that there are no major changes for this version when changing from infinite to finite populations. Once again, these results differ from behavioral experiences (*e.g.* [1]) where investors and trustees often act cooperatively, thus, motivating adding complexity to the model in an attempt to better emulate human behavior.

Secondly, for the Trust Game version with reputation, as we can see in Fig. 5.3 and Fig. 5.4 and by the trust and trustworthiness results, the added complexity do the model (introduction of reputation) had a positive result regarding the promotion of cooperation in both investors and trustees, particularly in the latter. If we only consider two possible strategies, *i.e.* either cooperate or defect, in the behavioral experiment in [1], namely with the version with social history (which is in a way mimicked by introducing reputation), we can see that about 89% of investors and 71% of trustees cooperated. In our simulations, results were 22% of the players cooperated when playing as investors and 65% when playing as trustees (on average, over time), showing there is still a considerable difference, mainly in regard to trust, and thus motivating us to once again add complexity to the model, namely by structuring the population in a SF network. Lastly, one should note that doing this comparison is not completely fair since: firstly Berg et al. [1] uses a relatively small number of subjects; secondly, in our simulations, we only consider two possible strategies, a player either cooperates or defects, however, in [1] investors can choose the precise amount they want as an initial stake and trustees can do the same for the return amount.

5.3.2 Structured Populations

As in the previous section, here we divide the discussion in the three versions considered with structured populations.

5.3.2.A Asymmetric role assignment

As mentioned above, additionally to the asymmetric role assignment variation, in this version, 3 different values for the investors' initial stake multiplication factor are considered. Let us first start by analysing the effects of only structuring the population in an SF network, *i.e.* for a multiplication factor of 3 and a $\lambda = 0$ in equation 4.2. For these values the results were that on average, (over time) 50% of the players playing as investors and 46% playing as trustees acted cooperatively, meaning that trust levels increased considerably (more than double) and trustworthiness values decreased by 15%.

Regarding the multiplication factor, when its value increases, the game becomes ”easier”, since the payoff for cooperative trustees increases proportionally. As the multiplication factor increases so does the number of trustworthy trustees, which may be explained by the fact that the payoff difference between cooperating and defecting decreases. This happens regardless of the λ considered, however, for higher values of λ the trustworthiness results are considerably better (Fig. 5.6), *i.e.*, contrarily to the Uber's example we gave before when the players with the larger degree in the network act as investors with a higher probability it favors the promotion of trustworthiness. Furthermore, increasing these players' probability of playing as trustees seems to have a negative effect.

Lastly, one would think that for higher λ values the trust values would be higher as well since a larger number of cooperative trustees means investors will have a better payoff when adopting a cooperative strategy as well. Results for trust (Fig. 5.5), however, show that for lower multiplication factors, contrarily to intuition, results are the opposite of the trustworthiness results, *i.e.* players with larger degree acting as trustees with a higher probability (negative λ values) favors the promotion of trust. When the multiplication factor increases, however, having the larger degree players playing as investors with a higher probability increases the trust values, as we expected.

5.3.2.B Diversity in the reputation

For this version, there were not any major improvements in terms of promotion of trust and trustworthiness. However, there are still some differences, namely as we can see in Fig. 5.7 and Fig. 5.8 the variation where reputation is assigned while taking into account players' degrees (players with a higher degree have a higher reputation and players with a lower degree have a lower reputation than before) shows better results, in regard to both trust and trustworthiness values, than the other two.

One explanation for these differences would be that, even though we are keeping the average reputation in the network the same, the fact that we are assigning higher reputation values to the players with the larger degree increases the probability of these players cooperating when playing as trustees since now their payoff when interacting with cooperative and defective (by having higher reputation values, defective investors will still trust them with a significant amount) investors is considerably closer than before. Because these players have a larger degree, they also have a higher probability of being chosen in the imitation process, thus increasing the number of cooperative trustees. As a consequence, this increase may in turn influence positively the number of cooperative investors, since more cooperative trustees makes cooperating more profitable.

Lastly, one should note that the positive effects of this alternative distribution of reputation (2 different values of reputation instead of 1) only has a positive effect when the players' degrees are taking into account, since the same 2 values are used in the variation with a random distribution and results are extremely similar to the uniform variation (same reputation value to all the players).

5.3.2.C Hybrid societies with pathological players

This last version of the game was the one where we obtained the best results (if only considering the results for the regular multiplication factor of 3), namely for the variation where the pathological players are assigned whilst taking into account players' degree. Even though this is also the only version where we force certain players to always cooperate, the number of players chosen as pathological is always really small when compared to the population size, *e.g.* for a threshold of 10 only 29 players are selected and for a threshold of 40 this number decreases to only 2 players our of 500.

Despite the reduced number of pathological players selected, as we can see in Fig. 5.9 and in Fig. 5.10, when these players are chosen while taking into account their degree the trust and trustworthiness values increase considerably, particularly for the first two thresholds considered. This increase is rather noticeable not only when compared to previous versions, but also when compared to the variation with the exact same number of pathological players chosen randomly, for which essentially results are the same regardless of the number of pathological players selected.

In regard to the evolutionary dynamics of the population, for the regular version of the Trust Game, with reputation, played in a network (Fig. 5.11), the dynamics are similar to the unstructured version with reputation. For the version with pathological players Fig. 5.12 (considering the lowest degree threshold), however, the most likely direction of evolution of the population is extremely different, namely, the population tends to evolve to a state where almost all the investors and about 80% of the Trustees have a cooperative strategy, showing that the introduction of pathological players in the population, when taking into account their degree during their assignment, has an extremely positive effect concerning the promotion of trust and trustworthiness.

One explanation for this major increase in the number of cooperative players of both roles would be the fact that the players with the larger degree, as mentioned before, have a higher probability of being chosen in the imitation process, furthermore because they are also linked to more players they have the potential to have higher payoffs making them more likely to be imitated by other players. Due to both these reasons, and by forcing these players to always cooperate we are increasing the chances of other players (non pathological players) adopting a cooperative strategy as well.

6

Conclusion

Contents

In this chapter, we firstly present some conclusive notes, secondly a summary of our contributions and lastly we introduce some potential future work in this area.

In this thesis we focused on studying the evolution of human trust and trustworthiness. In order to do this, we use evolutionary game theory and the Trust Game by Berg et al. [1] as our main tools. We used, as a starting point, the behavioral results of the original experience with the Trust Game [1], as well as several replications of this one, most of them included in the trust games meta-analysis by Johnson & Mislin [8]. Results from these experiments show that a considerable amount of investors makes the initial transfer and many trustees return a substantial quantity as well, *i.e.*, as explained in chapter 5, high values of trust and trustworthiness were observed in the behavioral data. Employing Game Theory (namely by considering the original payoff matrix), however, leads to rather different results, since in the unique Nash Equilibrium for this game the investors transfer zero money [1].

With the objective of more accurately modeling human behavior in the trust game, various studies were done where researchers add complexity to the original model in different ways. Some of the more important (and the ones we mainly focused on) were: considering altered payoff matrices to account for reputations, evolutionary game theory, and different populations structures. Thus, the main question that motivated this work was "which mechanisms explain the promotion of trust and trustworthiness in the context of the trust game?". In the first chapters of this Thesis, we analyse a group of works where these and some other variations are studied. In chapter 4 we develop our own model where we apply the variations mentioned concurrently, followed by some results of computational simulations in 5.

6.1 Contributions

In this section, we state the main contributions of our work as well as a summary of the main conclusions we drew from the results of our experiments. These results are detailed in chapter 5. In summary, they are:

Study of the dynamics in finite populations: We propose a new model (detailed in chapter 4) where we use the Trust Game payoff matrix with reputations, introduced by Sigmund in [2], and apply it to a 2-player version of the game with finite populations, using evolutionary game theory as a framework.

Study of the effects of structuring the population: We firstly studied our model when applied to unstructured populations, although other works using different models, as in [3], already consider reputation in an unstructured population context. Secondly, we introduce a new version where we consider reputation in a population structured in a scale-free network, allowing us to study some new components exclusive to populations structured in networks, listed below.

Study of the effects of introducing asymmetric role assignment: In the context of SF networks we studied the effects of forcing a dependence between a player role and his degree, controlled by equation 4.2, while varying the difficulty of the game. We concluded that the setup that is most favorable for the promotion of trust is: for a more difficult game having the higher degree players act as trustees; for a lower difficulty game having the higher degree players act as investors. Concerning trustworthiness, we concluded that having the higher degree players play as investors, regardless of the game difficulty, is always the most favorable.

Study of the effects of introducing diversity in the reputation: We studied the effects of different distributions of the reputations values in the population: a version where all the players have the same reputation value; a version where the reputation values are randomly distributed; a version where players with a higher and lower degree have a higher or lower reputation value respectively. We concluded that the distribution where we take into account the players' degrees is the one that most favors the promotion of both trust and trustworthiness.

Study of the effects of introducing pathological players: We studied the effects of introducing pathological players in the population, *i.e.* players that always cooperate regardless of the role they are playing as. Two different versions with pathological players were considered, one where they were selected randomly and another where the pathological players correspond to the players with a higher degree. We concluded that by selecting the higher degree players to be pathological players the values for both trust and trustworthiness increase considerably.

6.2 Future Work

Throughout this thesis, namely in the related work chapter, there are some ideas that were not implemented, due to time reasons or because we chose another approach on the problem. Below we list the ones we thought most relevant.

Network topologies: In the part of our work where we use structured populations we only consider a specific SF network. It could be beneficial to consider our model, not only on other network topologies (*e.g.* Erdős-Rényi networks), but also on larger SF networks and with different average degrees.

Update rules: In our simulations in this Thesis, concerning the imitation process, we consistently consider the same update rule. As it is done in [5] with their model, it would be of interest to study the effects of different update rules in the evolution of the population, namely the consequences it would have in regard to the promotion of trust and trustworthiness.

Types of reputation: In our model, we consider a type of reputation that only affects trustees when these act cooperatively, *i.e.* a variable v in the payoff matrix (Table 4.1) that corresponds to the probability cooperator trustee becomes known. This type of reputation is firstly introduced in [2], however, later in the same book a new type of reputation is proposed that we believe may yield interesting results as well. This second type of reputation affects defective trustees by punishing them, *i.e.* a new variable in the payoff matrix that corresponds to the probability that a trustee with a defective strategy becomes known.

Dynamic reputation: Furthermore, a different type of reputation system that could provide interesting insights would be for every player to have a dynamic reputation, *i.e.* each player would have a new reputation variable associated with him that would increase or decrease every time he acted cooperatively or defectively, respectively.

Bibliography

[1] J. Berg, J. Dickhaut, and K. McCabe, "Trust, reciprocity, and social history," *Games and Economic Behavior*, vol. 10, no. 1, pp. 122 – 142, 1995.

[2] K. Sigmund, *The calculus of selfishness*. Princeton University Press, 2010.

[3] M. L. Manapat, M. A. Nowak, and D. G. Rand, "Information, irrationality, and the evolution of trust," *Journal of Economic Behavior & Organization*, vol. 90, pp. S57 – S75, 2013.

[4] M. Chica, R. Chiong, M. Kirley, and H. Ishibuchi, "A networked$N$-player trust game and its evolutionary dynamics," *IEEE Transactions on Evolutionary Computation*, vol. 22, no. 6, pp. 866–878, 2018.

[5] M. Chica, R. Chiong, J. Ramasco, and H. A. Abbass, "Effects of update rules on networked n-player trust game dynamics," *CoRR*, vol. abs/1712.06875, 2017.

[6] M. Chica, R. Chiong, M. T. P. Adam, S. Damas, and T. Teubner, "An evolutionary trust game for the sharing economy," in *2017 IEEE Congress on Evolutionary Computation (CEC)*, 2017, pp. 2510–2517.

[7] J. F. Nash *et al.*, "Equilibrium points in n-person games," *Proceedings of the national academy of sciences*, vol. 36, no. 1, pp. 48–49, 1950.

[8] N. D. Johnson and A. A. Mislin, "Trust games: A meta-analysis," *Journal of Economic Psychology*, vol. 32, no. 5, pp. 865 – 889, 2011.

[9] M. Kandori, "Social norms and community enforcement," *The Review of Economic Studies*, vol. 59, no. 1, pp. 63–80, 1992.

[10] F. P. Santos, F. C. Santos, and J. M. Pacheco, "Social norms of cooperation in small-scale societies," *PLoS computational biology*, vol. 12, no. 1, p. e1004709, 2016.

[11] T. Wu, F. Fu, Y. Zhang, and L. Wang, "Adaptive role switching promotes fairness in networked ultimatum game," *Scientific reports*, vol. 3, p. 1550, 2013.

[12] A. Mao, L. Dworkin, S. Suri, and D. J. Watts, "Resilient cooperators stabilize long-run cooperation in the finitely repeated prisoner's dilemma," *Nature communications*, vol. 8, no. 1, pp. 1–10, 2017.

[13] J. M. Pacheco and F. C. Santos, "The messianic effect of pathological altruism," *Pathological Altruism*, p. 300, 2011.

[14] F. P. Santos, J. M. Pacheco, A. Paiva, and F. C. Santos, "Evolution of collective fairness in hybrid populations of humans and agents," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 6146–6153.

[15] M. J. Osborne *et al.*, *An introduction to game theory*. Oxford university press New York, 2004, vol. 3.

[16] J. M. Smith and G. R. Price, "The logic of animal conflict," *Nature*, vol. 246, no. 5427, p. 15, 1973.

[17] J. M. Smith, *Evolution and the Theory of Games*. Cambridge university press, 1982.

[18] J. Hofbauer and K. Sigmund, *Evolutionary games and population dynamics*. Cambridge university press, 1998.

[19] I. Bohnet and R. Zeckhauser, "Trust, risk and betrayal," *Journal of Economic Behavior & Organization*, vol. 55, no. 4, pp. 467 – 484, 2004.

[20] B. Kerr, M. A. Riley, M. W. Feldman, and B. J. Bohannan, "Local dispersal promotes biodiversity in a real-life game of rock–paper–scissors," *Nature*, vol. 418, no. 6894, p. 171, 2002.

[21] M. A. Nowak and R. M. May, "Evolutionary games and spatial chaos," *Nature*, vol. 359, no. 6398, p. 826, 1992.

[22] D. G. Rand, M. A. Nowak, J. H. Fowler, and N. A. Christakis, "Static network structure can stabilize human cooperation," *Proceedings of the National Academy of Sciences*, vol. 111, no. 48, pp. 17 093–17 098, 2014.

[23] F. C. Santos, M. D. Santos, and J. M. Pacheco, "Social diversity promotes the emergence of cooperation in public goods games," *Nature*, vol. 454, no. 7201, pp. 213–216, 2008.

[24] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.

[25] F. C. Santos, J. M. Pacheco, and T. Lenaerts, "Evolutionary dynamics of social dilemmas in structured heterogeneous populations," *Proceedings of the National Academy of Sciences*, vol. 103, no. 9, pp. 3490–3494, 2006.

[26] J. B. V. Huyck, R. C. Battalio, and M. F. Walters, “Commitment versus discretion in the peasant-dictator game,” *Games and Economic Behavior*, vol. 10, no. 1, pp. 143 – 170, 1995.

[27] H. Abbass, G. Greenwood, and E. Petraki, “The n-player trust game and its replicator dynamics,” *IEEE Transactions on Evolutionary Computation*, vol. 20, pp. 1–1, 2015.

[28] E. L. Glaeser, D. I. Laibson, J. A. Scheinkman, and C. L. Soutter, “Measuring trust*,” *The Quarterly Journal of Economics*, vol. 115, no. 3, pp. 811–846, 2000.

[29] D. Malhotra, “Trust and reciprocity decisions: The differing perspectives of trustors and trusted parties,” *Organizational Behavior and Human Decision Processes*, vol. 94, no. 2, pp. 61 – 73, 2004.

[30] E. Fehr, “On the economics and biology of trust,” *Journal of the European Economic Association*, vol. 7, pp. 235–266, 2009.

[31] A. S. C. T. Nowak, Martin A. and D. Fudenberg, “Emergence of cooperation and evolutionary stability in finite populations,” *Nature*, vol. 428(6983), pp. 646–650, 2004.

[32] P. Erdos and A. Rényi, “On the evolution of random graphs,” *Publ. Math. Inst. Hung. Acad. Sci*, vol. 5, no. 1, pp. 17–60, 1960.

[33] R. A. Holley and T. M. Liggett, “Ergodic theorems for weakly interacting infinite systems and the voter model,” *The annals of probability*, pp. 643–663, 1975.

[34] P. A. P. Moran, *The statistical process of evolutionary theory*. Clarendon Press, 1962.

[35] L. A. Imhof, D. Fudenberg, and M. A. Nowak, “Evolutionary cycles of cooperation and defection,” *Proceedings of the National Academy of Sciences*, vol. 102, no. 31, pp. 10 797–10 800, 2005.

[36] M. A. Nowak and K. Sigmund, “Evolution of indirect reciprocity,” *Nature*, vol. 437, pp. 1291–1298, 2005.

[37] ——, “Evolutionary dynamics of biological games,” *Science*, vol. 303, no. 5659, pp. 793–799, 2004.

[38] L. Rendell, R. Boyd, D. Cownden, M. Enquist, K. Eriksson, M. W. Feldman, L. Fogarty, S. Ghirlanda, T. Lillicrap, and K. N. Laland, “Why copy others? insights from the social learning strategies tournament,” *Science*, vol. 328, no. 5975, pp. 208–213, 2010.

[39] A. Traulsen, M. A. Nowak, and J. M. Pacheco, “Stochastic dynamics of invasion and fixation,” *Phys. Rev. E*, vol. 74, p. 011909, Jul 2006.

[40] F. C. Santos and J. M. Pacheco, “Scale-free networks provide a unifying framework for the emergence of cooperation,” *Physical Review Letters*, vol. 95, no. 9, p. 098104, 2005.

[41] F. C. Santos, J. Rodrigues, and J. Pacheco, “Graph topology plays a determinant role in the evolution of cooperation,” *Proceedings of the Royal Society B: Biological Sciences*, vol. 273, no. 1582, pp. 51–55, 2005.