Computational Models of Trust

Thesis submitted in accordance with the requirements of
the University of Liverpool for the degree of Doctor in Philosophy
by
Elisabetta Erriquez

March 2012

To my family

Abstract

Trust and reputation are key issues in the multi-agent systems domain. As in human societies, software agents must interact with other agents in settings where there is the possibility that they can be exploited. This suggests the need for theoretical and computational models of trust and reputation that can be used by software agents, and accordingly, much research has investigated this issue.

The first part of this thesis investigates the conjecture that agents who make decisions in scenarios where trust is important can benefit from the use of a social structure, representing the social relationships that exist between agents. To this end, we present techniques that can be used by agents to initially build and then progressively update such a structure in the light of experience. As the agents interact with other agents they gather information about interactions and relationships in order to build the network of agents and to better understand their social environment. We also show empirical evidence that a trust model enhanced with a social structure representation, used to gather additional information to select trustworthy agents for an agent's interactions, can improve the trust model's performance.

In the second part of this thesis, we concentrate on the context of coalition formation. Coalition stability is a crucial issue. Stability is the motivation of an agent's refusal to break from the original coalition and form a new one. Lack of trust in some of the coalition members could induce one agent to leave the coalition. Therefore we address the current model's limitation by introducing an abstract framework that allows agents to form distrust-free coalitions. Moreover we present measures to evaluate the trustworthiness of the agent with respect to the whole society or to a particular coalition. We also describe a way to combine the trust and distrust relationships to form coalitions which are still distrust-free.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

> *I turn up to give a lecture at* 9 *am on a Monday morning, trusting that my students will attend; and they in turn reluctantly drag themselves out of bed to attend, trusting that I will be there to give the lecture. When my wife tells me that she will collect our children from school, I expect to see the children at home that night safe and sound. Every month, I spend money, trusting that on the last Thursday of the month, my employer will deposit my salary in my bank account; and I trust my bank to safeguard this money, investing my savings prudently. Sometimes, of course, my trust is misplaced. Students don't turn up up to lectures; my bank makes loans to people who have no chance of repaying them, and as a consequence they go bankrupt, taking my savings with them. But despite such disappointments, our lives revolve around trust: we could hardly imagine society functioning without it.*
>
> Michael Wooldridge, Foreword from [21]

The sociologist Niklas Luhmann said *"A complete absence of trust would prevent one even getting up in the morning"* [44]. *Trust* is pervasive in human societies. It is necessary for many everyday tasks. Until the end of the nineties, trust was considered a notion appropriate only for human (or animal [14]) interactions and widely studied by sociologists and anthropologists. However, the creation of the Internet and the emergence of virtual societies has made trust a relevant research topic for Computer Scientists [22]. Computer Science has moved from the pattern of the isolated computer systems to the idea of a network of systems and distributed computing. This evolution has several implications for the security models, the policies and the mechanisms needed to protect users' information and resources in an increasingly interconnected computing infrastructure.

Similarly, Artificial Intelligence has moved to the idea of social, collective and independent intelligence, with the creation of the paradigm of agents and multi-agent systems [36]. For example, agents (computer systems)[1] can now participate in online auctions and negotiate on behalf of their owners. With this level of independence, agents need to be able to make decisions for themselves, including those regarding honoring their commitments. In this context, risks associated with mistrust and deception arise. Therefore, trust relationships need to be created in virtual societies as well as human ones.

In this chapter we give an introduction to multi-agent systems, provide some definitions of *trust* and describe the key contributions of this work.

1.1 Agents and Multi-Agent Systems

In this section, we provide some of the basic concepts of multi-agent systems, which will be used throughout this work. First, we give a definition of an intelligent autonomous agent:

> An *agent* is a computer system that is *situated* in some *environment*, and that is capable of *autonomous action* in this environment in order to meet its delegated objectives [70].

In this work, we assume that all the agents are *rational*. As defined in [61], *for each possible percept sequence, a rational agent should select an action that is*

[1] More details on agents will be provided in the next section.

expected to maximize its performance measure, given the evidence provided by the percept sequence and whatever built-in knowledge the agent has.

From the agent definition, a number of properties can be elaborated to better describe what capabilities are required for an agent to be *intelligent* [70]:

- **Reactivity**: Intelligent agents are able to perceive their environment, and respond in a timely fashion to changes that occur in it in order to satisfy their design objectives.

- **Proactiveness**: Intelligent agents are able to exhibit goal-directed behaviour by *taking the initiative* in order to satisfy their design objectives.

- **Social ability**: Intelligent agents are capable of interacting with other agents (and possibly humans) in order to satisfy their design objectives.

As the *social ability* property shows, it is clear that most of the problems delegated to agents require the interaction among multiple agents. A *multi-agent system* is a system containing multiple agents interacting together typically in order to solve problems that are beyond the capabilities of any individual agent [70]. In an open multi-agent system, agents can be owned by different organisations and may have common, complimentary or conflicting goals. Moreover, they can join or leave the community at any time. Therefore an open multi-agent system has the following properties [12]:

- **Dynamism:** Services provided in the society might become unavailable or the prices of products can change over time due to agents leaving the community or new agents joining it.

- **Openness:** As agents can join or leave any time, the number of agents in the society is not fixed.

- **Insecurity:** In the system, there may be incompetent, unreliable, or even malicious agents. Moreover, agents are self-interested, therefore, each agent is responsible for security, and does not rely on some global or special authority.

- **Distributed:** It is not practical for agents to rely on access to global system information, therefore no agent knows everything about the environment. Constant and completely reliable communication would be required to allow even one agent to obtain global information, and the computational costs required to create the global viewpoint often would exceed any performance improvements.

In such environments, trust can play a role as important as it assumes in human societies.

1.2 Definition of Trust

So far, we have related the notion of trust with scenarios that involve decision making in which the actions of different entities are relevant, but not known in advance. However, trust does not have a single unique accepted definition.

The Oxford English Dictionary defines trust as *confidence, strong belief, in the goodness, strength, reliability of something or somebody*. According to Marsh [45], most research on trust has come from three areas: sociology, psychology and philosophy. Marsh also describes four researchers as major contributors to this work: Morton Deutsch, Niklas Luhmann, Bernard Barber and Diego Gambetta.

Deutsch's definition [25] implies that trust is dependent on individuals and their perceived cost and benefits analysis of the given scenario, while Luhmann [44] argues that the concept of trust is a means of reducing complexity in society, when facing a decision. Barber [10] also links trust with society and cultures but not with an individual's perceived value of cost and benefits as in Deutsch's ideas. He links trust with expectation about the future.

The most interesting definition, from a computer science point of view, is Gambetta's [9]. He defined trust as follows:

> *Trust (or, symmetrically, distrust) is a particular level of the subjective probability with which an agent assesses that another agent or group of agents will perform a particular action, both before it can monitor such action (or independently of its capacity ever to be able to monitor*

> *it) and in a context in which it affects its own action. When we say we trust someone or that someone is trustworthy, we implicitly mean that the probability that he will perform an action that is beneficial or at least not detrimental to us is high enough for us to consider engaging in some form of cooperation with him.*
>
> *Correspondingly, when we say that someone is untrustworthy, we imply that the probability is low enough for us to refrain from doing so.*

With this definition, trust opens itself to be modelled mathematically and so it becomes more concrete than abstract compared to the other definitions. Moreover, this definition makes trust quantifiable. Trust is modelled as a probability, therefore it has a range from 0 to 1, where 0 is complete distrust and 1 is blind trust. Blind trust is an example of complete trust where one agent has complete trust in another regardless of the circumstances. Actions are dependent on this probability and those situations where trust in someone has no influence on our decisions are excluded.

This definition leads to the consideration that trust is relevant only when there is a possibility of deception. When someone is not completely trusted, there is a chance that the action he performs may be non-beneficial to us. This deception could be intentional, when actions are malicious but also caused by incompetence, carelessness or superficiality.

Moreover Gambetta's definition relates trust with cooperation or in general when any sort of interaction between agents occurs.

The lack of trust represents a real obstacle in many situations, from building personal relationships in human life to the establishment of on-line services or automatic transactions. Lack of trust also means having to make an effort and spend time to protect ourselves against deceit.

In contrast, trust promotes cooperation. It enables people to interact efficiently and similarly for companies and other organizations. Therefore a high level of trust is very attractive for the wealth of the community. On the other hand, the ability not to trust helps to avoid harm when faced with unreliable and unscrupulous persons.

Undoubtedly, trust has an important role in guiding people through uncertainty and risks.

The study of trust and reputation has many applications. Trust and reputation systems have been considered as key factors in the success of electronic commerce [7]. These systems are used as mechanisms to search for reliable partners but also to decide whether to honor contracts or not. They are also used as trust-enforcing, deterrent, and incentive mechanisms to avoid cheaters and frauds. But e-markets are not the only area of application. Trust has been used also as a way to improve belief revision mechanisms [11] or in cooperation and teamwork [71, 38][2] in multi-agent systems.

Gambetta's definition relates trust with cooperation, since cooperation requires some level of trust. In fact, if trust is only unilateral or there is complete distrust between agents then there cannot be any cooperation. A higher level of trust generally leads to a higher probability of cooperation.

It can be argued that blind trust can make cooperation successful since there is no possibility of distrust, but, it is important to note that blind trust could be an incentive to deceive. Consider the well-known game from game-theory *Prisoner's Dilemma* [8]:
In its classical form, the prisoner's dilemma (PD) is presented as follows (from wikipedia):

Two suspects are arrested by the police. The police have insufficient evidence for a conviction, and, having separated both prisoners, visit each of them to offer the same deal. If one confess and testifies (defects) *for the prosecution against the other and the other remains silent, the betrayer goes free and the silent accomplice receives the full* 10 *year sentence. If both remain silent* (cooperate)*, both prisoners are sentenced to only six months in jail for a minor charge. If each betrays the other, each receives a five-year sentence. Each prisoner must choose to betray the other or to remain silent. Each one is assured that the other would not know about*

[2] For a wider reading on the problem of learning cooperative strategies in competitive settings, see [47, 16].

the betrayal before the end of the investigation. How should the prisoners act?

In the game, regardless of what the opponent chooses, each prisoner always receives a lesser sentence by confessing and betraying the other prisoner. Suppose that Prisoner A thinks that Prisoner B will cooperate. This means that Prisoner A could cooperate as well, getting just the six months sentence, or he could defect, getting, in this case, the freedom. So it pays for Prisoner A to confess, therefore defect, if he thinks Prisoner B will cooperate staying silent. In the other case, suppose that Prisoner A thinks that Prisoner B will defect. Therefore, it is better for Prisoner A to defect as well, otherwise it would get a 10 years sentence.

Hence, it is better for Prisoner A to defect if he thinks Prisoner B will cooperate *AND* it is better for Prisoner A to defect if he thinks Prisoner B will defect. So, Prisoner A can accurately say, *No matter what Prisoner B does, I personally am better off confessing than staying silent. Therefore, for my own sake, I should defect*. However, if the other player thinks similarly, then they both defect and confess and both get a higher sentence (lower payoff) than they would get by both staying silent. Rational self-interested decisions result in each prisoner being worse off. They would both benefit by staying silent but they can't because if one stays silent the other one would benefit by confessing. Hence the dilemma.

The fundamental issue here is trust, as it is assumed that the two prisoners cannot communicate with each other. If one player could be sure the opponent chooses to remains silent (cooperates), which implied blind trust, then it would be its rational choice to defect leading it to go free. However, it is clear that they both could have got away with only six months each if they had been able to mutually trust each other to cooperate.

The prisoner's dilemma is an excellent example of how taking trust into account means a better result for the whole community and for the individual agent, but it also shows how blind trust could be exploited and therefore cooperation might not be successful.

Section 2 will provide a more detailed classification of various aspects of trust.

1.3 Overview and Main Contribution

The work in this thesis contributes to the trust research state of the art in two areas. Although the contributions in this thesis can be divided in two parts, they have a common target. The ultimate goal of the whole research is to allow agents to make better choices of the partners to interact with.

The first part of this work is related to the area of computational trust and reputation [3] models. We present a methodology to combine concepts of social networking and trust relationships to enhance a computational trust model oriented to those environments where social relations play an important role.

Social network analysis is the study of social relationships between individuals in a society, defined as a set of methods specifically developed to allow research and analysis of the relational sides of these structures [35]. In the multi-agent systems context, social structure analysis can play a vital role. The research in [66] asserts that the social ties created by social networks influence member behaviour through perceived social gains and rewards for themselves. In particular, social networks allow agents to strategically revise their relations. Similarly, our work tries to give better information for the agents on which to base their strategic decisions.

Although the idea of considering social relations has been previously studied, little attention has been given to the practicability of the models, neglecting to formulate how agents can gather information about these social relations. With this research, we attempt to overcome this limitation. The main characteristics of our contribution can be summarised as follows:

- Presenting a method for an agent to build a social network representation of their local environment;
- Using interaction information such as directed interaction or reputation information to update such representations;
- Providing an implementation of such representations of the social structure;

[3] A definition of reputation will be given in the next chapter

- Presenting empirical evidence that a technique to build and maintain a social network representation of the environment allows a trust model to be more effective in trustworthy partner selection.

As mentioned in the previous Section, trust and reputation have been considered key factors for the success of electronic commerce systems. For instance, reputation is widely used in electronic markets, such as for instance Ebay [60] and Amazon, as trust-enforcing, deterrent or incentive mechanisms to avoid cheaters and deception. The work carried out in this first part is presented in Chapters 3 and 4.

Another area of application in agent technology is teamwork and cooperation. In this area we can find extensive research about coalition formation techniques. The second part of this work concentrates on using trust as a factor in coalition formation models.

Cooperation is the foundation of multi-agent systems, allowing agents to interact to achieve their goals. Coalition stability is a crucial problem. If agents do not trust the other components of the coalition, they could break away from the alliance. Therefore, trust plays an important role for coalition stability.

Existing models of coalition formation use information about reputation and trust to rank agents according to their level of trustworthiness, therefore only considering a view of a single agent referred toward a particular target agent it wants to interact with. Hence, these models are inherently *local*: they lack a *global* view. In this part of our research, we address this limitation.

The main characteristics of the contribution regarding this part of our work can be summarised as follows:

- Offering a *global* view on trust and coalition formation;

- Providing a *formal* framework, ATF (Abstract Trust Framework) to consider trust in coalition formation;

- Providing a way to form *satisfying* coalitions for all the members, with regard to the trustworthiness of agents;

- Providing several notions of *mutually trusting* coalitions;
- Providing an actual illustration and analysis of the models proposed.

The work carried in this part is presented in Chapters 5, 6 and 7.

1.4 Publications

This thesis includes work that has been published in the following papers:

- Elisabetta Erriquez, Wiebe van der Hoek, Michael Wooldridge (2010) Building and Using Social Structures in Agent ART. In the 13th International Workshop on Trust in Agent Societies, May 10-14, 2010 Toronto, Canada.
- *(To appear)* Elisabetta Erriquez, Wiebe van der Hoek, Michael Wooldridge (2011) Building and Using Social Structures: A Case Study using the Agent ART Testbed. ACM Transactions on Intelligent Systems and Technology (TIST)
- Elisabetta Erriquez, Wiebe van der Hoek, Michael Wooldridge (2011) An abstract framework for reasoning about trust. In: The 10th International Conference on Autonomous Agents and MultiagentSystems (AAMAS-11). May 2-6, 2011 Taipei, Taiwan.
- Elisabetta Erriquez, Wiebe van der Hoek, Michael Wooldridge(2011) An abstract framework for reasoning about trust. In the 14th International Workshop on Trust in Agent Societies, May 2-6, 2011 Taipei, Taiwan.
- *(To appear)* Elisabetta Erriquez, Wiebe van der Hoek, Michael Wooldridge. A formal analysis of trust and distrust relationships in Shakespeare's Othello. In F. Paglieri, L. Tummolini, R. Falcone & M. Miceli (Eds.), The goals of cognition. Essays in honor of Cristiano Castelfranchi. London: College Publications.

The work presented in the first three papers is illustrated in Chapters 3 and 4, while the work presented in the last three papers is illustrated in Chapters 5, 6 and 7.

1.5 Structure of the Thesis

The remainder of this thesis is organized as follows:

- **Chapter 2** attempts to define trust and reputation, and reviews the state of art on the current approaches to modelling trust in agent systems and in coalition formation. It also summarises the open issues.

- **Chapter 3** presents a method for agents to build a social network representation of their local environment.

- **Chapter 4** describes the methodology and the test settings used to evaluate our approach. It also provides an analysis of the results.

- **Chapter 5** presents ATF, the abstract framework devised in this research to allow trustworthy coalition formation.

- **Chapter 6** presents an extension to our ATF model, allowing the use of both *trust and* distrust relations.

- **Chapter 7** uses Othello, the famous tragedy by Shakespeare, to give an illustration and analysis of the models proposed.

- **Chapter 8** concludes the thesis and outlines the directions for future work.

Chapter 2

Trust

Just as in human societies, sofware agents have their own objectives. The most rational strategy for an agent is to maximize its own return. Often, agents must interact with other agents in settings where there is the possibility that they can be exploited. This suggests the need for computational models of trust and reputation that can be used by software agents, and accordingly, much research has investigated this issue over the past decade.

This Chapter is structured as follows: Section 2.1 will present a classification of the various types of trust identified in the literature. Section 2.2 presents a review on the main computational approaches to model trust and reputation. Section 2.3 reports problems influencing the effectiveness of the trust models presented and highlights open issues that need to be addressed and the requirements for this research. Finally, Section 2.5 presents a survey of the state of art regarding coalition formation models which consider trust as a factor.

2.1 Trust Classification

On a general level, there are two main approaches to deal with trust [59]:

1. **Individual-level trust**, whereby an agent has some beliefs about the honesty or the potential behavior of its interaction partner;

2. **System-level trust**, in which the agents in the system are forced to be trustworthy by rules or protocols that regulate the system, like, for example, in auctions, voting, contract-nets or market mechanisms.

The first approach provides each agent with a model to reason about certain features of their counterpart in order to make decisions about their reliability, honesty and so on. This approach allows the agents to calculate a measure of trust they can put in the counterpart. A high degree of trust towards an agent means that the agent is likely to be chosen as an interaction partner. A low degree of trust means that the agent would probably not be selected or that the counterpart agent would adopt a different strategy to deal with an untrustworthy agent. Therefore, the purpose of the first approach is to lead the agent to decisions such as how and whom to interact with.

The second main approach is based on the design of mechanisms or norms that govern the interactions between agents. These mechanisms need to ensure that if the agents behave correctly then they will gain some utility whereas malicious agents will loose utility or be penalised. Therefore, these protocols must guarantee that the agents will have no better option than interacting honestly with each other.

In this thesis we focus on the first approach; for more details about the second approach see [59].

It is possible to classify trust models according to the information sources they use to calculate their trust values. Direct and indirect interactions are the most common source of information used in most computational models [40, 63, 67, 74, 45, 65]. Recently, attention has been given to information related to sociological aspect of agents' interactions. Using multiple sources of information can help to reduce uncertainty in different situations [57].

Direct Trust We refer to the value obtained from the analysis of direct interactions as *direct trust*. On a general level, direct interactions are, undoubtedly, the most reliable source of information for a trust model. All computational trust models use this type of information. This information is based on observations made by an agent a about its past interaction with a particular agent b. These observations are reflected in agent *a*'s opinion of agent *b*'s behaviours and used to predict how *b* will behave in future interactions.

In this category, we can also include information deriving from observable interactions between other agents. However, in most scenarios, this type of information is rarely available since interactions among other agents are not publicly visible.

Indirect Trust We refer to the value obtained from the analysis of information from indirect interactions (interactions between two other agents) as *indirect trust*. Indirect information, also called witness or reputation information or recommendations, is the information coming from other agents in the society. This information is normally based on the witnesses own direct interaction but it can also be based on other indirect interactions.

Indirect information is normally easy available, however it carries a degree of uncertainty that the information based on direct experience does not have. It is not uncommon that agents manipulate or hide information for their own benefit. For this reason, indirect information is not ready to use as direct information but requires additional processing in order to be used in computational trust models.

Reputation is built on observations of an agent's past behaviour. This includes evaluations collected from other agents in the society. Every agent has a personal point of view, and so the values may be recorded in different ways. Aggregating these values presents difficulties that are not examined in this thesis [1].

Direct trust and reputation have a close relationship. Direct trust in an agent is formed based on observations made by the agent after direct interactions. The reputation of an agent is constructed from observations of other agents about their own direct interactions (provided that the observations are reported accurately). Therefore, direct trust can be seen as reputation at a local level.

In multi-agent systems, reputation can be useful when there are a large number of agents interacting (for example, online auctions or stock-trading). Reputation enables buyers, for instance, to choose the best seller (based on ratings about previous interactions) in the system. Therefore, reputation can induce sellers to be well

[1] For more information, see [59].

behaved if they think that a low reputation value will cause buyers to avoid them.

Reputation is also a useful source of information to form an initial trust between agents who have not yet interacted directly.

Social Trust We refer to the value obtained from the analysis of social information about the agents' society as *social trust*.

Reputation can be considered a social concept, however there are several other social relationships that affect an agent's reputation. For example, belonging to a government office may imply high trustworthiness of the information the agent provides. Moreover, the kind of social interactions between agents can help to determine the trustworthiness of the reported information. Therefore, we can distinguish two types of social relationships: the type based on the role the agent has in the society and the ones whose type is inferred from the interactions between two agents. The first type is usually publicly known by all the agents and it reflects initial roles or membership of an agent in a particular organization, for example, company employee. The second type emerges during interaction between agents, reflecting a certain type of relationship, for example a supplier or frequent buyer [40].

Another source of information proposed by some researchers [40] is a trusted third party, a sort of authority that supplies information about agents when required. This information is considered to be honest and reliable. However, in the majority of scenarios, this kind of information is rarely available.

Social information is important to allow the agent form a richer view of its environment so that it is more likely to make good decisions in terms of whom to trust and whom to ask for recommendations.

2.2 Computational Trust and Reputation Models

In recent years, many different computational trust and reputation models have been presented, all with their own characteristics. In this section we present a selection of these models in order to provide an overview of the area.

One of the earliest computational trust model is the one proposed by Marsh [45]. The model takes into account only direct interactions. It distinguishes three types of trust:

1. ***Basic trust***, which models the general trusting disposition independently of whom the agent is interacting with. It is calculated from all the experiences accumulated by the agent. Good experiences lead to a greater disposition to trust, and vice versa.

2. ***General trust***, which is the trust that one agent has toward another without considering any specific situation. It simply represents general trust of the other agent.

3. ***Situational trust***, which is the trust that one agent has in another one taking into account a specific situation. The utility, the importance of the situation and the *General trust* contribute to calculate the *Situational trust*.

In order to define the overall value, the author proposes three methods: *the mean, the maximum and the minimum*. Also, three different types of agents are defined: the optimistic (which uses the maximum trust value from the range of experiences it has had), the pessimistic (which uses the minimum trust value), and the realistic (which uses a value calculated as the average over a certain set of situations). These trust values are used to help an agent to decide whether to cooperate or not with another agent. Besides trust, the decision mechanism also considers the importance of the action to be performed, the risk associated with the situation, and the perceived competence of the target agent. To calculate the risk and the perceived competence, the three types of trust (basic, general, and situational) are used. In this way, the model tries to bring into the calculation other factors affecting the decision of the agent to trust his partner.

Marsh's model is considered the first prominent, comprehensive, formal, computational model of trust. His intent was to address an imperfect understanding, a plethora of definitions, and informal use in the literature and in everyday life with regard to trust. However, his model is often simplified due to the difficulty of finding values for some variables used to compute trust (for example, importance, utility, competence, risk, etc.).

Online reputation models, such as the ones used by Ebay [60] or Amazon are based on the ratings that users perform after the completion of a transaction. The user can give three possible values: positive (1), negative (-1) or neutral (0). The reputation value is computed as the sum or the mean of those rating over a certain period of time.

The Sporas system [74] extends the reputation model of Ebay using a new formula to calculate the amount of change of the reputation value according to the new rating value. Sporas only considers the most recent rating between two users. Users with very high reputation values perceive smaller rating changes after each update than users with a low reputation. To take into account the reliability of such values, Sporas uses a measure of reliability based on the standard deviation of reputation values.

Models of this type are based on a centralised approach and are not suitable for open multi-agent systems, where there is no central authority.

The Histos model [74] is an enhancement to Sporas that takes into account the group dynamics as in Regret [63]. In particular, Histos looks at the links between users (in a social network) to deduce personalised reputation values, to deal with the lack of personalization of reputation values that Sporas had. This enables an agent to assemble ratings from those it trusts already rather than those it does not know. The ratings are represented as a directed graph where nodes represent agents and edges carry information on the most recent reputation rating given by one agent to another. The root node represents the agent owner of the graph. The reputation of an agent at level x of the graph is calculated recursively as a weighted mean of the rating values that agents in level $x-1$ gave to that agent. The weights are the reputation values of the agents that rate the target agent. The agents who have been rated directly by the owner of the graph have a reputation value equal to the rating value. This is the base case of the recursion. The model also limits the length and number of paths that are taken into account for the calculation. However, the reputation value does not depend on the context and no special mechanisms are provided to deal with cheaters.

Both Histos and Sporas have been shown to be robust to collusion. This is because ratings from those agents that are badly rated themselves have a diminished effect on the reputation of others and those they might want to protect. However, as the authors point out themselves, the major drawback is that users are reluctant to give bad ratings to their trading partners. This is because the model is not incentive-compatible, and so agents have no incentive to give ratings in the first place.

The model proposed by Yu and Singh [19] tries to deal with absence of information. The main contribution of their work is the aggregation of information obtained from referrals while coping with a lack of information. The model is based on witness information and has a mechanism to locate information sources based on the individual agent's knowledge. Each agent in the system has a list of agents it knows and their associated expertise. When the agent needs information, it can send a query to a number of agents in its list, who will try to answer the query if they can. In case they cannot, the agents will send back a list of other agents who they believe could answer the query. This constitutes the referral system.

Yu and Singh use the Dempster Shafer theory of evidence to model information the agent retrieves [73]. When an agent does not receive any rating (good or bad), Yu and Singh's model considers this a lack of belief (or disbelief) therefore a state of uncertainty (where all beliefs have an equal probability of being true). Dempster's rule allows beliefs obtained from various sources to be combined to support the evidence that a particular agent is trustworthy or not. Moreover, the agent holds a *local* belief about the trustworthiness of another related to its direct interaction with that particular agent. In these cases, ratings obtained from witnesses are discarded. However, Yu and Singh do not deal with the possibility that an agent may lie about its rating of another agent. They assume all witnesses are totally trustworthy even if an agent could obtain some benefit by lying about their rating of an opponent.

With regard to the problem of deceitful witnesses, Schillo et al. [65] propose a possible solution. They propose a Prisoner's Dilemma set of games with a partner selection phase. Each agent receives the results of the game it has played plus the information about the games played by a subset of all players (its neighbors). As a result, the agent can have an impression of the partner's honesty based on the behavior they exhibited according to the normal Prisoner's Dilemma actions (cooperation or defection). An interesting feature of this model is the use of a data structure called *TrustNet* to store third party information to complement the results from the initial phase. This structure is similar to the graph used by Histos [74]. In fact, the *TrustNet* is a directed graph where nodes represent agents and edges carry information about the observation that the parent node has referred to the owner of the TrustNet about the child node.

While Yu and Singh's model shows the effectiveness of referrals and Dempster Shafer's theory in modelling reputation, Schillo's model illustrates how witness information can be used to reason effectively against lying agents. These models, however, although make use a *social network*, they simplify direct interactions and fail to frame such interactions within the social setting with regards to the type of relationships that exist between the witnesses and the potential interaction partners.

To overcome this limitation, Sabater and Sierra [63] adopt a sociological approach closer to real-life scenarios within their model called Regret. Their reputation value, which represents the trust to be placed in the counterpart, is a weighted sum of subjective opinion derived from direct interactions, the group opinion of the opponent and the agent's opinion of the opponent's group. These last two opinions compose the social dimension of reputation. Although Regret takes into account the possibility of dishonest reports, it assumes that agents are willing to share their opinions about one another. This model is interesting because it tries to frame the interactions between agents within the social setting, using social relations to overcome the lack of information that the agents may face. Regret uses the social network to find groups of witnesses, to decide which witnesses will be consulted, and how to weigh those witnesses' opinions. Regret also uses the concepts of neighborhood reputation and system reputation, which are calculated using fuzzy rules from the reputation of the target's neighbor agents. Finally, the system also makes use of information regarding the social role of the agent in the society, for example seller, buyer, etc.

However, this is useful only if the information based on the particular domain is available. In fact, Sabater and Sierra assume that this information is always available to the agent. In Regret, each agent owns a series of so called sociograms, a representations of the social network in the form of a graph, which describes relationships between all the agents in the society. One of the limitations of the Regret model is the fact that it is not shown how each agent can build the sociograms that they use in their trust computations.

The Fire model [40] tries to overcome this limitation using a referral system similar to the one used by Yu and Singh. The model integrates a number of information sources to produce a comprehensive assessment of an agent's likely performance in open systems. Specifically, in order to provide trust metrics, Fire uses four measures:

- interaction trust, or direct trust, as defined in section 2.1, which is a trust value based on previous direct interaction with the agent it wants to assess;
- role-based trust, which is a value based on the particular context the agent is operating in;
- witness reputation, or indirect trust, as defined in section 2, which is a trust value based on information from other agents;
- and certified reputation, a reliable trust value provided by a third trusted part, if available.

These four measures are calculated by different modules in the system, and combined together to produce a comprehensive assessment. The modularity of the system allows the agent to be able to produce a trust value even in absence of information from a particular module. For instance, if agent *a* has not yet interacted directly with agent *b*, the Fire system could still produce a trust assessment based on the measure for the other modules. A key part of this model is the use of *references* (endorsements of trust from other agents), which are used when no reputation or other sources of trust exist. This feature enables Fire to provide a trust metric in cases where other models fail to produce an assessment because of lack of information about a particular agent.

However, one limitation of the model is the assumption that all the agents will be always willing to share information and that they will never lie in doing so. In real scenarios, this assumption might be too strong. Moreover, Fire does not use the social context as the Regret model does. Fire does not evaluate interactions within the context of the social relationships existing between all the parts involved, for example, the witness agents, the target agent and the agent itself.

2.3 Issues of Trust Models

After the analysis of the trust and reputation models presented in the previous subsection, there are some considerations to be made.

Most approaches use a mathematical-based design for their trust model and some also inherit concepts from game theory. The main sources of information used are the direct interactions and information from third party agents from indirect interactions. There are only few models [63, 40, 30] using other aspects of the interactions to enrich the values of trust and reputations calculated. We agree with Sabater and Sierra [63] that a good mechanism to increase the efficiency of trust and reputation models is the introduction of sociological aspects as part of these models. It is this aspect of interactions between agents that we concentrate on in the first part of this thesis.

Beside the general issues of calculating trust and reputations measures, there are other issues that can affect the performance of a trust model. These comprise the correlated evidence problem and the dynamism of open multi-agent systems.

The *correlated evidence* problem happens when the opinions of different witnesses are based on the same event(s) or when there is a considerable amount of shared information that tends to unify the witnesses' way of thinking. In both cases, the trust on the information shouldn't be as high as the number of similar opinions may suggest [63]. Sabater and Sierra use graph theoretic techniques to deal with this problem. In this thesis, we used the same approach with a simplification on the number of the sociograms used. More details are given in Section 3.

In an open multi-agent system, the number or behaviour of agents can change at any time. New agents can enter the system and old agents can leave it at any time. This means that an agent needs to be able to continuously learn about new agents and be aware that a previous interaction partner may have left the environment. Moreover, relationships among agents may change due to changes in agents' behaviour or role. Therefore, a previously trusted agent might become a liar or the quality of its service might degrade due to different circumstances.

Given the possible dynamism of an open multi-agent system, a trust model should be able to adapt itself to cope under these changes and function reasonably consistently.

Regret [63] uses a social network to take into account different types of relationships among agents and it also does not impose any constraints on the dynamism of the system. However, as mentioned in the previous paragraph, it is assumed that each agent already owns the sociograms used in the trust computations and it is not shown how the agents can build or update the sociograms should any interaction change the relationships among the agents. This means that agents have no tools to update their network representation in cases such as an agent breaking a trust relationship, or new agents appearing in the system.

In the first part of this thesis, we present techniques for agents to build a social network representation of their local environment and maintain their own representation of such environments in case of changes. With this extended perception of the environment, agents can make more informed decisions.

2.4 Testbed

In this paragraph we describe the Agent ART (Agent Reputation and Trust) testbed [34], which has been used for the evaluation of the methodology proposed in this thesis.

The Agent ART Testbed initiative was launched with the goal of establishing a testbed for agent reputation and trust related technologies. Although no longer maintained, the ART Testbed is designed to serve in two roles:

- as a competition forum in which researchers can compare their technologies against objective metrics, and

- as an experimental tool, with flexible parameters, allowing researchers to

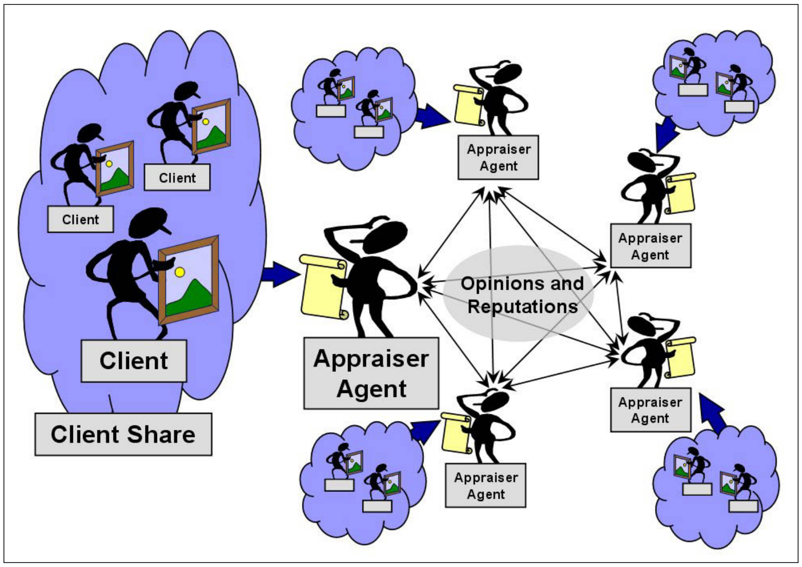

Figure 2.1: Game overview in the Agent ART Testbed from [33]

perform customizable, easily-repeatable experiments.

The Agent ART Testbed claims to be a versatile, universal experimentation platform, which looks into relevant trust research problems and unites researchers toward solutions via unified experimentation methods. The testbed provides metrics and tools that researchers can use for comparing and validating their approaches. The testbed also serves as an objective means of presenting technology features, both advantages and disadvantages, to the community.

A game in the Agent ART testbed is intended to support the comparison of different strategies of agents as they act in combination. We decided to use the Agent ART testbed as an experimental platform because it covers relevant trust research problems and helps researchers find a unified solution. Moreover it provides metrics through which it is possible to compare and validate different trust models.

The context of the testbed is the art appraisal domain, where agents function as painting appraisers with varying levels of expertise in different artistic eras. Figure 2.1 shows an overview of a game in the testbed. Clients request appraisals for paintings from different eras; if an appraising agent does not have the expertise to complete the appraisal, it can request opinions from other appraiser agents. Appraisers receive more clients, and thus more profit, for producing more accurate appraisals.

In what follows we provide a description of how a game in the Agent ART Testbed is structured, as described in [33].

In each timestep, multiple clients present each appraiser (agent) with paintings to be appraised, paying a fixed fee f for each appraisal request. To increase business, appraisers attempt to evaluate paintings as closely to market value as possible. A given painting may belong to any of a finite set of eras (a painting's era is known by all appraisers), and appraisers have varying levels of expertise in each era. An appraiser's expertise, defined as its ability to generate an "opinion" about the value of a painting, is described by a normal distribution of the error between the appraiser's opinion and the true painting value. The agents do not know each other's expertise values.

Appraiser agents may also purchase from each other reputation information about other appraisers. Appraiser agents must decide when, and from whom, to request opinions and reputation information to generate accurate appraisals for clients.

The simulation engine oversees each operation, including client requests, appraiser opinion generation, transactions between appraisers, and returning final appraisals to clients.

When an agent wishes to request an opinion about a particular painting, it sends a request message to another agent (potential opinion provider), identifying the painting to be appraised. Upon receiving an opinion request, if the potential provider is willing to provide the requested opinion, it responds by sending a certainty assessment about the opinion it can provide, defined as a real number between zero and one (one represents complete asserted certainty), based on its expertise. The potential provider is not required to provide a truthful certainty assessment. If the potential provider does not wish to participate in the requested transaction, it may choose to decline the request. By sending a certainty assessment, the provider promises to deliver the requested opinion should the certainty assessment be accepted by the requester. After receiving the provider's certainty assessment, the requester either sends payment to the provider if it chooses to accept the promised opinion, or sends a "decline" message if it chooses not to continue the transaction. The cost of each transaction is a non-negotiable amount. Upon receipt of payment, the provider is not required to send its actual opinion, neither is the provider forced to send any opinion at all.

Upon paying providers, but before receiving opinions from providers, the requesting appraiser is required to submit to the simulation engine its roster of opinion providers and a set of corresponding weights. Weights are values between zero and one, loosely representing the appraiser's confidence or trust in each provider's opinion. The final appraisal requested is calculated by the simulation engine as a weighted average of eligible opinions, using the weights communicated by the agent to weigh each opinion. At the end of the transaction, the true value of the painting is revealed to the agent by the simulation engine. Having learned the final appraisal and the painting's true value, the agent may use feedback to revise its trust model of other appraisers.

In addition to conducting opinion transactions, appraisers can exchange reputations, or information about the trustworthiness of other appraisers. The protocol follows the same payment procedure set for the opinion transaction. Although a reputation value has the same form as the weights mentioned before, appraisers are not required to report to requesting appraisers the same weights submitted to the simulation. Although these weights represent the providing appraiser's subjective trust measures, a requesting appraiser can learn how to interpret the provider's weights after observing the relationships among several weights sent by the same provider over time.

Each appraiser has a bank account, monitored by the simulator, from which it pays transaction costs and into which are deposited client appraisal fees. The winning appraiser agent is selected as the appraiser with the highest bank account balance at the end.

Since, in the Agent ART testbed, asking for reputation information has a cost, we believe that using the social structure to select only agents that are more likely to provide meaningful information could translate to a cost saving. In our model, once the information is gathered, the opinion from the witness agent selected is weighed taking into account the social relationships linking the agent with the witnesses and the target agent.

2.5 Trust in Coalition Formation

Many tasks cannot be completed by a single agent, often because of limited resources or capabilities of the agents. Sometimes even if a task could be completed by a single agent, the performance of the single agent might be not acceptable. In such situations, agents may form groups to solve the problem by cooperating. Many organizational paradigms have been distinguished in the context of MAS (see [39]). One of these paradigms is coalition formation.

Coalition formation is a fundamental form of interaction that allows the creation of coherent groups of autonomous agents in order to efficiently achieve their individual or collective goals. Forming effective coalitions is a major research challenge in the field of multi-agent systems. Central to this endeavour is the problem of determining which of the possible coalitions to form in order to achieve some goal.

The main issues in coalition formations are

- Coalition structure generation;
- Teamwork;
- Dividing the benefits of cooperation.

It is beyond the scope of this work to elaborate on these issues (for more information see [64]).

As mentioned in the previous subsection, in open distributed systems, there are many components that can enter and leave the system as they wish. Therefore, the notion of *trust* becomes key when it comes to decisions about which coalitions to form, and when. When such systems are inhabited by agents that encompass some degree of autonomy, each representing their own stakeholders with their own objectives, it not only becomes plausible that some agents are not trustable, the consequences of joining a coalition of which some members cannot be trusted, or do not trust each other, becomes a key aspect in the decision of whether or not to join a group of agents.

In the next subsection we present a survey of the state of art regarding coalition formation models that consider trust as a factor.

2.5.1 Related Work in Trust in Coalition Formation

Cooperation inherently involves a certain degree of risk, arising from the uncertainty associated with interacting with self-interested agents. Trust has been recognised as a critical factor in assessing how likely it is for an agent to fulfill his commitments [45, 20]. Normally, agents infer trust values based on experience regarding previous interactions.

Our primary aim in the second part of this work is to consider trust in the context of coalition formation. To date, there are very few coalition formation models that consider trust as a factor. Core [56] is a trust model that tries to solve the agent coalition formation problem. The model characterizes an agent's trust from two aspects: *competence* and *reputation*. Competence is used to establish the extent to which an agent is fit for a particular job. Agents possess a series of competences, which in Core are expressed as a vector space. To carry out each task requires certain competences, and these requirements are also expressed as a vector space. Euclidean distance is used to find the best fitting agent for a particular task according to his competences. The appropriate agents for forming a coalition are then selected according to their competence fitness and their trust values.

In [38], the concept of a *clan* is introduced. A clan is a group of agents who trust each other and have similar objectives. In [38], coalitions are assumed to have a long-term life, although often still directed towards a particular goal. Therefore task-based approaches typically result in cooperation for a single goal, requiring a team to be created for subsequent goals even if the team members are the same.

Clans are considered to have a “medium-term” life and motivation is used as a factor in the decision making process of whether to form a clan or accept the request of joining a clan.

In both models, the coalition or clan formation process has to be initiated by an individual agent. Therefore, the coalition formation process starts when an agent *i* asks other agents he deems trustworthy to join its coalition. When asked to join the coalition, an agent *j* can refuse the invitation if he does not trust *i* or he does not have motivation to accept.

In these models, we believe one important element is neglected. The models only consider the trust binding two agents *i* and *j*. Therefore if, for example a third agent *k* is also asked to join the coalition, he will only check if his level of trust in *i*, completely neglecting his opinion of *j*. We believe that this is an important issue that could damage the stability of a coalition.

The approach proposed in this thesis tries to overcome this problem, providing a rich and robust framework for agents to reason about the trust relationships among them in the community. We also propose practical ways to allow the formation of *distrust-free* coalitions.

2.5.2 Summary

In this chapter we have provided a classification of the various types of trust mentioned in the literature. We have provided an overview of the main computational approaches to model trust and reputation.We also have reported problems that current approaches have yet to tackle, which we try to address in this work.

Finally we have presented a survey of the state of art regarding coalition formation models which include trust as a factor.

Chapter 3

Building and Using Social Structures

3.1 Motivation

One important question in trust and reputation research is what sources of information agents can use to build their trust of others upon. As introduced in Section 2, agent *a* can base its trust or reputation of agent *b* using experience of previous interactions between them; or agent *a* might ask a third party *c* about its opinion regarding *b*. An important additional source of trust is to use information about the social relationship between agents [63]. We call the network of social relationships *social structure*. If *a* and *b* are competing for the same resources, for example, this may negatively affect the way they trust each other. Similarly, if agents *a* and *b* are likely to have complementary resources, and their cooperation would benefit both, it seems likely that they would be more inclined to trust each other.

Although models of social structure have begun to be considered in models of trust and reputation [63], to date, *implementing* social structures, and hence properly *evaluating* their added value and *validating* them, has not been done. And, most importantly, the issue of how a social structure *evolves* does not appear to have been considered in the literature. These are the issues we address in this work. We argue that agents who make decisions in scenarios where trust is important can benefit from the use of a social structure, representing the social relationships that exist between agents.

In this Chapter, we present a technique for agents to build a social network representation of their local environment. We also show empirical evidence that a trust model enhanced with a social structure representation which can be used by the agent to gather additional information to select trustworthy agents for its interactions, can improve the trust model's performance.

In order to compare different models for trust in agent communities we used the Agent ART testbed [34]. The Agent ART testbed is a platform for researchers in trust to benchmark their specific trust models. More details regarding the Agent ART testbed were given in section 2.4.

To illustrate our approach, we have taken two of the agents developed to compete in previous Agent ART testbed competitions and enhanced them with a representation of a social structure, and algorithms to build and refine this social structure. Then, we ran a number of competitions in which we evaluated and compared the performance of our *social* agents. In particular, we measured the number of successful interactions for each of them, the utility gained, and the number of rounds won by them in a competition. When evaluated against all of those measures, the *social* agents performed better than their asocial counterparts, which is encouraging with regard to exploring social structures further, and to extend experiments to other scenarios.

3.2 Social network analysis

Social network analysis is the study of social relationships between individuals in a society, defined as a set of methods specifically developed to allow research and analysis of the relational sides of these structures [35]. These social networks are represented using graphs called *sociograms*. A different sociogram is normally built for each type of social relationship to be examined, for example *kinship* or *friendship*. According to the relationship considered, the graph can be directed or undirected, with or without weighted edges. However, for our purpose, we will use a simplification of these multiple sociograms. In this work, we use a single sociogram that we call a *social structure*. Using multiple sociograms allows the agents to be linked to others in more than one social relationship at the same time. Conversely, using a single sociogram allows the agents to be linked together in just one social relationship at any given time. In Section 3.3.1, we explain in more detail how the social structure is created and how the different types of social relationships are taken into account.

In the multi-agent systems context, social structure analysis can play a vital role. For example, the search for relevant information involves finding the right sources – the agents who have the required information or expertise. Thus, social network analysis can be an important tool in discovering these relevant information sources. The interactions between individuals in the society can lead to the addition of new links in the social structure.

Social network analysis has also been used in market trading context. In [23] it has been considered how 'networks of influence' between market traders impact upon the agents' performance.

Agents are aware of the presence of other agents in the society because of their direct interactions, using services or asking opinions. However, interactions and recommendations are also useful in predicting the relationships among agents. An agent can infer from an indirect recommendation that a witness agent has used the target agent, (i.e., the agent the recommendation is about), as a service provider. This is because it is assumed that if two agents have interacted in the past, they will have an opinion about each other.

It might be considered debatable, but in practice, we surely use this kind of "social logic" every day. Asking for a recommendation about a mechanic or restaurant employs exactly this type of reasoning, taking into account the trust we have in the person we are asking and his trust of the mechanic or restaurant [37].

Similarly, these trust relationships are used by computational agents to minimise the uncertainty the agents have while interacting. Although direct interactions are perhaps the most reliable source of information to compute trust values, one cannot assume that each agent, at any stage, has interacted with every agent that is relevant. Therefore, an agent might not be able to form an opinion, based just on direct experiences, about every agent in the society. That is why "reputation" information from third party agents, called witnesses, can complement information gathered from direct interaction.

Although the aim of a recommendation from a third agent is to evaluate the trustworthiness of the target, the agent that is asking the recommendation is also able to draw a link, in its social structure, between the witness agent and the target agent, thus building a more complete view of its environment. This simple intuition is used in our approach to build the agent's social structure.

As mentioned in chapter 2, the goal of partner selection in social networks has been studied for a long time. In fact, Quattrociocchi et al [57] shown that using different sources of information, when there is no correlations among the different sources, contribute to uncertainty reduction in situations where trust is required. This is exactly what we aim to show in our work, but we concentrate on the specific contribution given by information from social relationships. Our tests aim to show that a social structure as a source of information, together with the common sources, can help agents in making more informed decisions in uncertain situations.

In fact, the ultimate goal of our work is to allow agents to make better choices of the partners to interact with.

3.3 Building the Social Structure

3.3.1 Identifying the Social Relationships

The concept of using a social structure to enhance a trust model was initially proposed in a computational agent in the Regret model [63]. In Regret, an agent owns a set of sociograms that show the social relations in the society. These sociograms are not necessarily complete or accurate. However, Regret does not propose a way for the agent to build sociograms.

The main contribution of this work is to show how the social structure can be built by each agent in the society and show, via evaluation, how using the social structure in a trust model can improve the performance of the trust model. Using the Agent ART testbed, we are able to use agents from past competitions and enhance their internal trust models with the social structure. As the agents interact with other agents they gather information about interactions and relationships in order to build the network of agents and to better understand their social environment.

As in Regret [63], we consider three main types of social relationship

- Competition (COMP). This is the type of relation found between two agents that pursue the same goals and need the same (usually scarce) resources. This is the kind of relation that could appear between two sellers that sell the same product or between two buyers that need the same product.
- Cooperation (COOP). This relation implies the exchange of sincere information between the agents and some kind of predisposition to help each other if possible. In other words, we assume that two agents cannot have at the same time a competitive and a cooperative relation.
- Trade (TRD). This type of relation is compatible either with cooperative or competitive relation. It reflects the existence of some commercial transactions between two agents, without a distinctive competitive or collaborative behaviour.

To identify the type of relationship between two agents in our Agent ART testbed, we use the concept of expertise. The Agent ART testbed assigns to agents different levels of expertise for each art era. The agents, during the game, can ask each other information about their levels of expertise. The agents are providers of a service which is the evaluation of a painting. However, they are, at the same time, consumers of this service, when their level of expertise is not sufficient to evaluate a particular painting. This means that agents with similar expertise compete in the society for the same market share. Agents with different expertise are more likely to cooperate because they will require each other's services.

The Agent ART testbed scenario is designed so that agents are globally in competition for client share, but able to cooperate on single appraisals. They are providers of a service, appraising a painting of a particular era, however they might need to buy that same service from other agents who are more expert than them in that era.

We now present a formal definition of the social structure.

Let $G = (V,E)$ be a graph where V is a set of vertices of G and $E \subseteq V \times V$ the set of edges of G. Formally, the social structure is defined as a undirected weighted graph $G = (V,E)$ where

- V is the finite set of vertices that represent the agents in the society,
- $E \subseteq V \times V$, is the set of the edges that correspond to links between agents, and
- each $(a,b) \in E$ has associated with it a weight w_{ab}.

The weights associated on each edge will be used to identify the social relationship between the two agents linked by that particular edge. To define this weight, we introduce the concept of *expertise distance*. To compute how similar or different the levels of expertise between two agents are, we use Euclidean distance. Each era is considered as a dimension in an n-dimensional Euclidean Space, where n is the number of eras. The total number of eras is defined in the games setting of Agent ART testbed. The values of the different levels of expertise identify points in the Euclidean Space. We can, therefore, define an agent x as a vector $v = a_1, a_2, \ldots, a_n$ where a_i is the level of expertise for era i, $i \in 1, \ldots, n$ and n, which is the total number of different eras, is set in the game configuration settings.

The Euclidean distance $d(a,b)$ between two n-dimensional vectors, or agents, is given by:

$$d(a,b) = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \cdots + (a_n - b_n)^2}$$

This distance is called the *expertise distance* and it is the weight assigned to the edge linking the agents a and b.

In the case where the agent does not have information about the expertise of the other agent in a particular era, it is assumed the agents have the same expertise.

If the agent assumes that the unknown value is at the low end of the range of expertise's value, this might artificially increase the distance measure. Setting it to the same value will be equivalent to not considering that dimension in the distance measure.

Agents who are a small distance apart have similar expertise and are considered to be in a social relation of *competition*. The further apart the agents are in the space, the more they depend and cooperate with each other, finding themselves in a social *cooperation* relationship. When the distance between agents is neither small nor large, we consider them to be in a *trade* relationship. This relationship does not indicate an unclear situation in which the agent does not know what to do, but it simply means that the two agents, according to their expertise, could occasionally cooperate or compete. In the evaluation, we use a singleton value, as parameter of the experiments, to assess the relationship, called *expertise threshold*.

3.3.2 Basic Concepts from Graph Theory

The Social Structure is represented as an undirected graph. Therefore, in this Section we present some basic notions of Graph Theory used later on [1].

Let $G = (V,E)$ be a graph, as introduced in the previous subsection. G is a directed graph iff E is a set of directed edges (otherwise it is called undirected graph). Note that it is equivalent to considering an undirected graph as defined by a symmetric relation E on $V \times V$.

A connected component of an undirected graph is a subgraph in which any two vertices are connected to each other by paths, and to which no more vertices or edges (from the larger graph) can be added while preserving its connectivity. This is known as a maximal connected subgraph [69].

A *dense* graph is a graph in which the number of edges is close to the maximal number of edges. The opposite, a graph with only a few edges, is a *sparse* graph.

Sometimes, a connected graph G can be so sparse that the removal of a single vertex will disconnect it. Such vertices are quite important in our work. They are called *cutpoints*. A vertex x is called a *cutpoint* in G if $G \setminus \{x\}$ contains more

[1] For a more complete and detailed background on graph theory see [69]

components than G does; in particular if G is connected, then a cutpoint is a vertex x such that $G \setminus \{x\}$ is disconnected [69]. Figure 3.1 shows an example.

From a sociological point of view, a cut-point can be seen as indicating some kind of local centrality (LC). In our context, it can be considered as the agent centralizing the largest amount of information among the agents in the components it connects. A similar concept is expressed from the central points. There are various measures of the centrality of a vertex within a graph that determine the relative importance of a vertex within the graph. We use the *degree centrality*, defined as the number of edges incident upon a vertex. Figure 3.2 shows an example. The degree is often interpreted in terms of the immediate risk of a node being passed through by whatever is flowing through the network, such as a virus, or in our case, information.

3.3.3 Building and Maintaining the Social Structure

The steps that agents follow to build their social structure are as follows: initially, the agent x starts with an empty social structure. Every time x directly interacts with another agent y, asking for a painting appraisal, an edge is created between x and y. Every time the agent x receives reputation information from a witness agent y about a target agent z, an edge is created between y and z. This is because, as mentioned before, it is assumed that an agent can infer from an indirect recommendation, or reputation information, that the witness agent has used the target as a service provider. The edges linking the agent that owns the social structure and other agents are a result of direct interactions. The edges linking the other agents in the social structure are the result of indirect interaction, or recommendation.

Algorithm 1 describes in pseudo code the procedure used to build the social structure.

As explained earlier, the use of the social structure in the trust model is inspired by Regret [63]. However the aim of our experiments is to analyse the impact of the social structure on a general trust model. Therefore the trust model of the agents selected is not changed to conform to the Regret trust model. Hence, the

Algorithm 1: An algorithm for computing the Social Structure *Soc* for agent *x*.

```
begin
    $Soc \leftarrow \emptyset$;
    while game is not over do
        // if x asks y for a painting appraisal

        if $askedAppraisal(x,y)$ then
            // an edge between x and y is added
               if not already present
            if $(x,y) \notin Soc$ then
                $Soc \leftarrow Soc + (x,y)$;
            // the weight of the edge is updated
            $weight[(x,y)] \leftarrow computeDistance(x,y)$;
        // if x asks y for information about z
        if $askedRepInfo(x,y,z)$ then
            // an edge between y and z is added  if not
               already present
            if $(y,z) \notin Soc$ then
                $Soc \leftarrow Soc + (y,z)$;
            // the weight of the edge is updated
            $weight[(y,z)] \leftarrow computeDistance(y,z)$;
    return Soc
end
```

task of dealing with potentially false information is left to the original trust model, according to its own strategy.

The agent trust model is integrated with the use of the social structure as follows. We use the social structure to find witnesses to ask for reputation information about other agents, to decide which witnesses will be consulted, and how to evaluate those witnesses' opinions.

The first step to calculate a witness reputation is to identify the set of witnesses (W). The initial set of potential witnesses is the set of all agents that have interacted with the target agent in the past. Starting from these agents, the social structure is filtered, leaving the portion that contains the selected agents. The result of this first step is a subgraph of the initial social structure. This graph can be formed by more than one connected component.

In a large society, the set of witnesses can be very big, and agents with frequent interactions are likely to have a considerable amount of shared information that tends to unify their way of thinking [53].

Algorithm 3 shows, in pseudo code, the steps required to compute the set of witnesses for a particular target t.

The agents in the final set of witnesses, W, will be asked for reputation information about the target agent.

According to Figures 3.1 and 3.2, the set W would be formed by agents *e, x, y* and *l*. If W or the initial set are empty then the agent will behave in the same way as without the use of the Social Structure. This situation can happen, for instance, at the beginning of the game, where a sufficient number of interactions has not taken place yet to allow the social structure to have meaningful information. The social relationship linking the target agent and the witness agent affects the way the reputation information is considered. For example, the agent is likely to discard reputation information from witness agents who are in strong competition with the target agent.

At every timestep, the social structure, and hence the weights on the edges, are updated according to the new information received from the witness agents, as illustrated in algorithm 1. This includes adding new links between the agents or modifying the weights in relation to any expertise change, as shown in algorithm 2. This helps the social structure to be dynamic and to reflect changes in behavior from the agents in the society.

We have integrated this strategy in two agents that were previously developed for the Agent ART Competition. From an initial examination of the implementation and documentation of the agents, we have observed that many of the agents participating in the Agent ART Competition did not implement the reputation module, available in the Agent ART testbed framework, in their trust model. They only relied on direct interactions, because the total population in the competition was not large, allowing them to interact directly with almost every other agent. The only agents implementing the reputation module were Simplet and Connected.

Simplet's trust model is inspired by Liar [48], whose model was introduced in Section 2.

Connected's trust model creates and updates at every timestep a list of "friends". These friends are selected by calculating the accuracy of the previous paintings' evaluations. Connected also has internal checks and different thresholds to decide how accurate the evaluation is.

The next Chapter provides a description of the framework used for our test and of the process used, together with an analysis of the results of our experiments.

Algorithm 2: An algorithm for assigning weight on the edge between agents *x* and *y*.

```
begin
    foreach $era \in Eras$ do
        // Gather information about expertise of agent $x$
           from the Agent ART testbed Simulation Engine
        $expertise_x[era] \leftarrow expertise(era, x)$;
        // Gather information about expertise of agent $y$
           from the Agent ART testbed Simulation Engine
        $expertise_y[era] \leftarrow expertise(era, y)$;
        // if we do not know both values, we assume them
           to be zero
        / if $expertise_x[era]$ and $expertise_y[era]$ is null then
            $expertise_x[era] \leftarrow 0$;
            $expertise_y[era] \leftarrow 0$;
        else
        // if we do not know one of the values for that
           era, it is assumed the agents have the same
           expertise
        if $expertise_x[era]$ is null then
            $expertise_x[era] \leftarrow expertise_y[era]$;
        else if $expertise_y[era]$ is null then
        $expertise_y[era] \leftarrow expertise_x[era]$;
          // Compute the Euclidean Distance between the
           two vectors of expertises
        $w \leftarrow computeEuclideanDistance(expertise_x, expertise_y)$;
        return $w$;
end
```

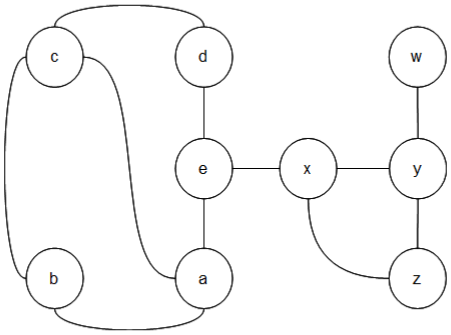

Figure 3.1: Cutpoints: e, x and y

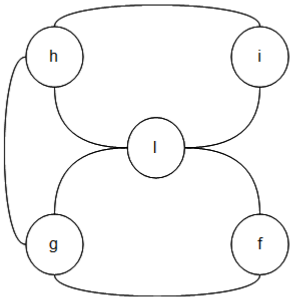

Figure 3.2: Central point: l

Algorithm 3: An algorithm for computing the set of witnesses *W*, from the Social Structure *Soc*, for target agent *t*.

```
begin
    $Ws \leftarrow Soc$;
    // Remove from $W$ vertices that are not linked to
       $t$
    $remove-all(W,\ x,\ y)$ where $x \neq t$ and $y \neq t$;
    // Identify the connected components of $W$
    $Comp(Ws) \leftarrow connectedComponents(Ws)$;
    // For each component, finds cut-points or
       central points
     foreach $W_i \in Comp(Ws)$ do  if $hasCutpoints(W_i)$ then
         $W \leftarrow W + cutPoints(W_i)$
    else  $W \leftarrow W + centralPoints(W_i)$   return $W$;
end
```

Chapter 4

Evaluation Methodology

One of the contributions of this work is to provide a systematic evaluation of the use of the social structure, implemented as explained in the previous section, in a trust model.

Therefore in this Chapter we provide a description of the framework used for our tests and explain the process used for the evaluation, the parameters and the metrics used. We also provide an analysis of the results.

4.1 Experiments and Analysis

Thirteen different configurations were developed. The parameters for each configurations are shown in Table 4.1. The configurations contain parameters for the simulation environment, the Agent ART testbed, and for the configuration of the social structure in SocialSimplet and SocialConnected. The parameters of interest for these tests are only a subset of those available for the simulation environment. We provide now a description of the parameters tuned in our experiments:

- *# Games*: this parameter states how many runs are carried out for each configuration;

- *# Timesteps*: this parameter states how many timesteps compose each game;

- *Certainty cost*: this parameter specifies the fixed fee paid by opinion requesters to opinion providers for a certainty;

- *Reputation Cost*: this parameter specifies the fixed fee paid by reputation requesters to reputation providers;

- *# Era to change*: this parameter states for how many eras the expertise value will be changed at each timestep;

- *Amount of expertise change*: this parameter specifies the amount by which each changing era will be modified;

- *# Instance of each agent*: this parameter specifies if more than one instance of the same agent will be competing in the game;

- *Initial reputation discard*: this parameter is proper to the social structure use. It decide whether the "social" agents will discard reputation information provided at the very beginning of the game, when it is assumed that the agents cannot have interacted with each other just yet.

- *Expertise threshold*: this value is used by the "social" agents when using the social structure, to decide what social relationships links the two agents examined in that particular instance.

For each configuration, 50 runs were carried out, as shown from the parameter *#games*. All configurations have a constant set of agents consisting of those used by the 2008 Agent ART Competition, (excluding Agent Uno [49]), plus SocialSimplet and SocialConnected. In this phase of our experiments, although the agents are allowed to lie, we tried to reproduce a fair competition environment. Therefore, excluding Agent Uno was necessary because its trust model uses knowledge of the design of the Agent ART testbed framework to tune the parameters of its trust model. These parameters are available to programmers because they are described in the documentation of the Agent ART testbed, however information regarding these parameters is not available to the agents during the competition, unless the designer explicitly incorporated it during the coding phase. Therefore, in our opinion, even if the parameters are available to designers, they should be deliberately ignored for the sake of not biasing the trust model.

The configurations can be grouped into three subcategories. The first category includes configurations A, B and C. This first group reflects the test setting used in the official Agent ART Competition[1]. The aim of this first group of configurations is to test the performance of the agents in settings with different dynamism.

Parameters such as *#era to change* and *amount of expertise change* are used to create different levels of dynamism. As mentioned above, the era represents the period of time a particular painting belongs to, and the expertise value represents how confident the agent is about its knowledge in a particular era. Therefore, by changing the level of expertise the agent has to react to variation in the other agents' responses to an opinion request which might be due to different expertise rather than strategy.

The second category includes configurations D, E, F, G, J. In this group we keep the value of dynamism for the parameters *#era to change* and *amount of expertise change* still to the intermediate value. The aim of this group of configurations is to test, in particular, the "social" agents with respect to their performance in a setting where the cost of communication among agents is removed. Since the agents using the social structure make heavy use of reputation and expertise[2] requests, it is possible that their final performance is affected by these costs. In this group of configurations, we also vary the total population, going from 1 instance of each agent in the game to 10 instances for each agent. We also want to test if the social structure proves to be more useful in a larger society and over a longer period of time, therefore in configuration J, the number of timesteps is raised to 200.

The final category is comprised of configurations H, I, L, M and N. Again, in this last group we consider the intermediate value of dynamism for the parameters *#era to change* and *amount of expertise change*. We also keep the cost of the com-

[1] From the Agent ART Testbed Web site. http://megatron.iiia.csic.es/art-testbed/competition2008.htm

[2] The expertise requests are called in the Agent ART Competition setting *certainty request* because the agent can communicate its level of certainty about an opinion on a particular era

munication among agents to zero. In this category, we test the performance of the agent by varying parameters of the social structure (bottom Section of Table 4.1), such as the *expertise threshold* used to identify the social relationship between two agents in the social structure. This set of configurations also helps to fine tune performance linked to these parameters in the Agent ART testbed.

The dark gray shaded cells in Table 4.1 identify the important parameters whose values were changed for the given group of configurations.

Parameters	Configurations												
	A	B	C	D	E	F	G	J	H	I	L	M	N
# games	50	50	50	50	50	50	50	50	50	50	50	50	50
# timesteps	90	90	90	90	90	90	90	200	90	90	90	90	90
# eras	20	20	20	20	20	20	20	20	20	20	20	20	20
avg client/agent	20	20	20	20	20	20	20	20	20	20	20	20	20
client fee	100	100	100	100	100	100	100	100	100	100	100	100	100
opinion cost	10	10	10	10	10	10	10	10	10	10	10	10	10
certainty cost	1	1	1	0	0	0	0	0	0	0	0	0	0
reputation cost	0.1	0.1	0.1	0	0	0	0	0	0	0	0	0	0
# opinion msg	3	3	3	3	3	3	3	3	3	3	3	3	3
# certainty msg	10	10	10	10	10	10	10	10	10	10	10	10	10
deny self opinion	true	true	true	true	true	true	true	true	true	true	true	true	true
# eras to change	1	3	3	3	3	3	3	3	3	3	3	3	3
amount expertise change	0.05	0.1	0.3	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
# instance of each agent	1	1	1	1	2	5	10	10	1	1	1	1	1

initial reputation discard	true	true	true	true	true	true	true	true	false	true	true	true	true
expertise threshold	15	15	15	15	15	15	15	15	15	2	5	10	20

Table 4.1: Configurations settings for the evaluation: Top section contains settings for the Agent ART testbed and the bottom section contains the settings concerning the social structure of SocialSimplet. Dark grey cells contain parameters changed for that configuration.

Several metrics are used to evaluate performance. Each metric is observed for Simplet and Connected and their counterparts using the social structure, which we call SocialSimplet and SocialConnected. The values are used to evaluate and compare the performance of the agent with its social copy using the social structure. The metrics considered are:

- the total utility, which is the bank balance at the end of the competition, as in the Agent ART Competition;
- the number of games won, grouped by configuration setting; and
- the number of fulfilled and violated interactions, compared with the total number of interactions.

Every *interaction* is intended as an exchange of information between agents. This information can be an opinion about a painting or reputation information about another agent. These two types of information can be considered as the two services provided by the agents in the environment. This final metric is particularly important in assessing the value of the social structure because it reveals how many times the trust model has been successful in selecting a trustworthy agent. To assess whether an interaction has been fulfilled or violated, the internal trust model of each agent is analysed. In both Simplet and Connected, the agents use the difference between the true value of the painting and the opinion provided by other agents in previous interactions. The true value of each painting is revealed after every interaction, allowing the agent to calculate this value. The difference in the values, together with the expertise of the agent providing the opinion in the considered painting's era, allows the agent to decide if the interaction was fulfilled or violated. This evaluation method depends on the agent internal trust model. Therefore if the difference is above a certain threshold, the interaction is considered violated. Again, this threshold may be different in Simplet and Connected because it is set in the internal trust model of each agent. The purpose of these experiments is not to compare the trust model of Simplet vs Connected, but to compare each of these agents with their social counterpart. Thus this threshold is not considered as a parameter of the experiments. Since each agent and its social copy have the same trust model, this threshold will have the same value in both agents, meaning the comparison is not affected by this.

4.2 Results

In Simplet, in Figure 4.1 we can observe the percentage of fulfilled interaction averaged over the fifty runs for each configuration. Although our calculations show that, overall, there is an improvement of only 3.64%, the total number of interactions is considerably larger in SocialSimplet, as shown in Figure 4.3.

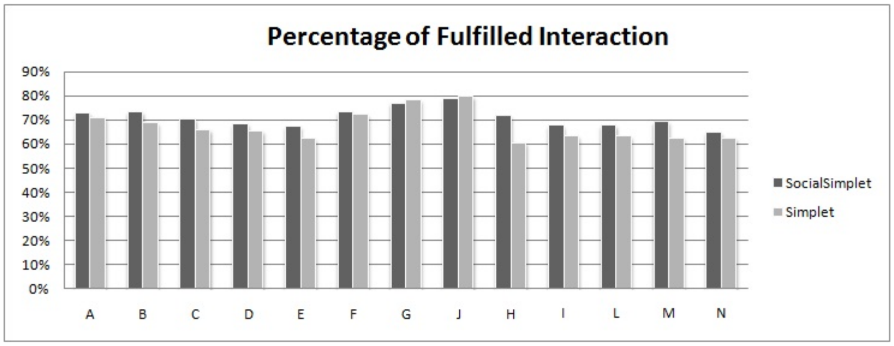

Figure 4.1: Percentage of Fulfilled Interaction for all configurations for Simplet and SocialSimplet.

The results show that SocialSimplet has, on average, 8200 more interactions than Simplet. Thus the effective number of fulfilled interactions is substantially larger. This improvement in the number of fulfilled interactions clearly shows that the internal trust model of the agent benefits from the information derived from the social structure. This suggests that the trust model with the social structure can more accurately assess which agent is more likely to be trustworthy.

Since the social structure is updated after every time step, SocialSimplet is only marginally affected by the change of dynamism in the configurations A, B and C, going from a 72.6% of fulfilled interactions in configuration A, to a 70.5% in configuration C. Simplet seems to suffer slightly more because of this change, losing a

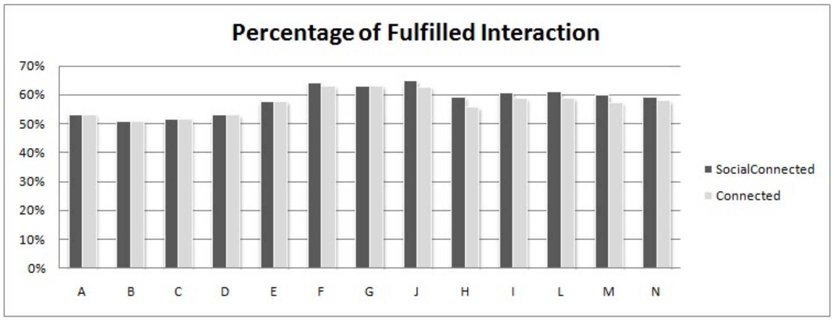

Figure 4.2: Percentage of Fulfilled Interaction for all configurations for Connected and SocialConnected.

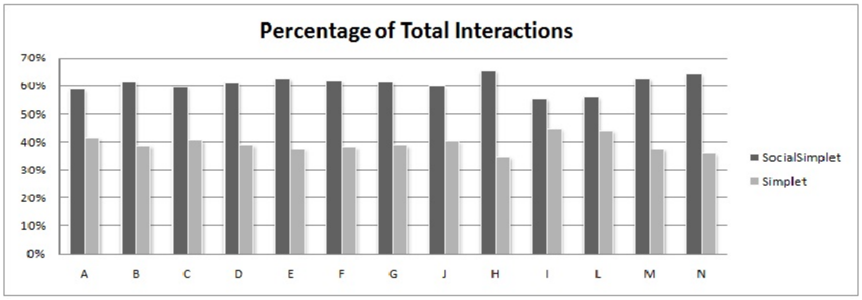

Figure 4.3: Percentage of Total Interaction for all configurations for Simplet and SocialSimplet.

little more than 5% over the three sets of games. A similar trend may be observed in Connected. Figure 4.2 shows an improvement of nearly 2%. Although this value is smaller than the percentage reached in Simplet's case, in this case there is a very large increase in the number of interactions by SocialConnected, with respect to the number of interactions in Connected, as shown in Figure 4.4, with over 350, 000 more interactions, on average, over all configurations. We have analysed the reason for this behaviour. We observed that the trust model of Connected comprises a series of constrains used to select to interact with. Therefore, it happens, at times,

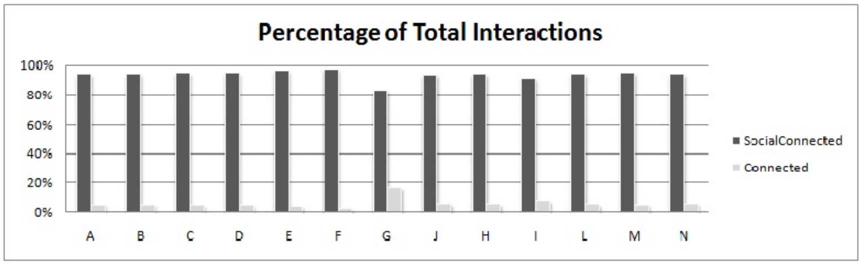

Figure 4.4: Percentage of Total Interaction for all configurations for Connected and SocialConnected

that the agent using Connected model ends up not asking for much information, because no agents satisfy its complex constraints. However, in SocialConnected, where the social structure is used for selecting the more trustworthy agents, most of the time, the agent manages to select a set of agents to interact with. This results in more interactions. In conclusion, the difference in the number of fulfilled interaction for SocialConnected is, on average, more than 2 million in each configuration. Also, SocialConnected is only marginally affected by the change of dynamism in the configurations A, B and C, going from a 53% of fulfilled interactions in configuration A, to 51.66% in configuration C.

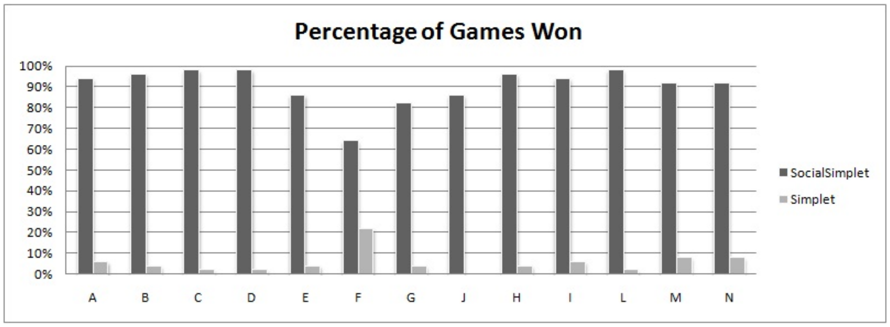

Figure 4.5: Percentage of Games Won for all configurations for Simplet and SocialSimplet.

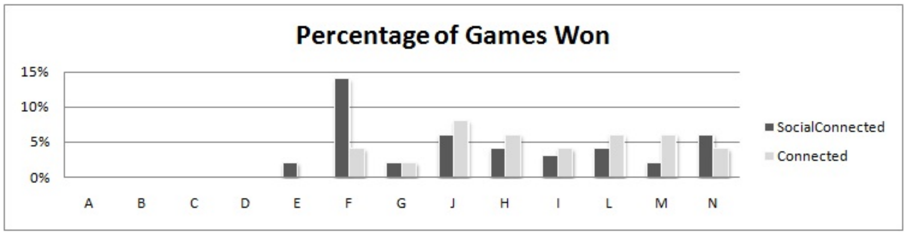

Figure 4.6: Percentage of Games Won for all configurations for Connected and SocialConnected.

Configurations F and G and J show to be amongst the highest percentage of fulfilled interactions among the other configurations, in both SocialSimplet and SocialConnected. This means that the social structure is particularly useful in those societies with a large number of agents, where direct interactions are not always possible with every agent in the society. In configuration J, SocialSimplet nearly achieves a total of the 80% of fulfilled interactions against a total of more than 130 thousand interactions, hence having more than $102,600$ successful interactions. SocialConnected reaches 65%, having more than 1 million successful interactions. This shows that the accuracy of the social structure improves over time, because after every interaction the social structure gains more information about the society. Therefore in very long games, such as the runs in configuration J, the agent can refine the social structure and obtain a better perception of the environment it inhabits.

In configuration H, we consider that, at the beginning of the game, agents cannot yet have meaningful information about the other agents in the society. Therefore we do not add links between agents in the social structure when we receive reputation information at the beginning of the game, if the agent thinks they are initial default values. In Figure 4.1 and Figure 4.2, we can observe that ignoring this initial information does not seem to adversely affect the performance of SocialSimplet nor SocialConnected. This is because, since the social structure is updated at every time step, the agent can correct any wrong assumption. From the set of games in configurations I, L, M and N, we can learn that a good expertise threshold value seems to be one in the range 10.081 to 15.081. The expertise threshold is used to assess the types of social relationships linking the agents. We can also note that in configurations D, E, F and G, where the reputation and expertise requests have no cost, there is a progressive improvement in the utility gained, in line with the increasing total population, in both SocialSimplet and SocialConnected.

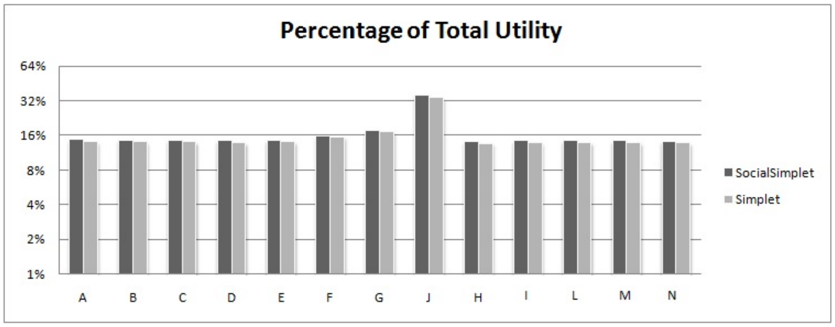

Figure 4.7: Percentage of Total Utility for all configurations for Simplet and SocialSimplet.

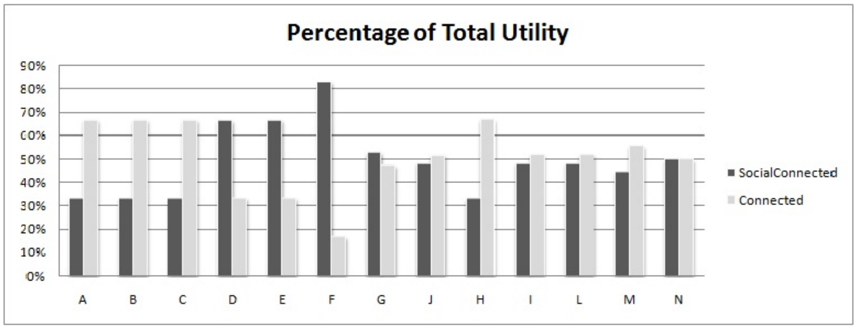

Figure 4.8: Percentage of Total Utility for all configurations for Connected and SocialConnected.

On the other hand, when considering total utility, SocialSimplet performs better over all the different configurations, but only by a small margin. Figure 4.5 and Figure 4.7 show, respectively, the percentage of games in which SocialSimplet wins, and the margin by which it wins. This shows that SocialSimplet wins very often (almost 100% of the games), but by a small margin. This is due to the client allocation share strategy, which is affected by the accuracy of the final appraisals produced. The agent chooses a total amount, representing the time taken to examine the painting, to be spent in generating its own opinion about a painting's value. This amount affects the accuracy of the generated final appraisal; this parameter is a strategic decision on how the agent manages its money, and it is not affected by the changes made in SocialSimplet. In other words, the way Simplet administers its money is not part of the trust model, even if it is affected by it to an extent. More details about this issue are presented in Section 4.3. The margin of improvement in the accuracy of the appraisals produced is due to the improved selections of other agents who are asked for their opinions, but it is harder to appraise given that the total utility is determined largely by the client allocation share strategy, which is the same for Simplet and SocialSimplet in this evaluation. Considering that the only difference in the two models, Simplet and SocialSimplet, is the use of the social structure, it seems reasonable to infer that any improvement in the performance is due to the social structure.

In Connected and SocialConnected we see similar behaviour. The difference in the total utility gained, between the two agents, is more evident, both in the case where Connected earns more and vice versa. We can observe, in Figure 4.8, that SocialConnected, gains much more utility when the costs linked to the formation of the social structure are removed, as shown in configuration D to N. This means that the agent has more money available to invest in the generation of a more accurate appraisal, hence affecting the client allocation share strategy and giving Social Connected the possibility, in the next time step, to improve its performance.

4.3 Remarks on the Agent ART Testbed

The results presented in the previous Section demonstrate that the use of a social structure can offer some benefits. However some remarks concerning the Agent ART testbed are necessary to put these benefits in their proper context.

During each run in the competition, the agent has a set of clients, and each client will pay a fixed amount to obtain an appraisal for one painting. Each painting is classified in one of the finite set of eras. For a given era, the expertise of an agent is modeled by an error distribution. To be more precise, an agent's appraisal is generated by the simulator, from a normal distribution whose mean is the painting's true value and whose deviation is determined by the agent's expertise and the money it spends.[3] Therefore, given that the expertise of the agent is predetermined by the simulator, an agent can only directly affect the accuracy of its appraisals by changing the amount of money it spends to generate the appraisals.

Although the client pays a fixed amount for the appraisals, the client's share allocation strategy depends on the accuracy of such appraisals. This means that low accuracy appraisals can affect the number of clients the agent gets and so, indirectly, its total utility.

Also as the agent has no way to control how many clients it can interact with it can't directly ”work harder” to improve its total performance.

Tuning these parameters correctly, we believe, is out of the scope of this work and also out of scope for a trust model. This is the reason why we felt it necessary to introduce, as measure of a successful trust model, the metric regarding fulfilled and violated interactions. This also explains why the success of the “social” agents are not proportionally reflected in the total utility.

4.4 Summary

In this chapter we have outlined a way to combine concepts of social networking and trust relationships. For the first time, this work has presented empirical evidence that a technique to build and maintain a social network representation of the environment allows a trust model to be more effective in selecting trustworthy agents. Agents use their social network to obtain knowledge that they could not gather otherwise, and use this knowledge to filter their trust relationships. Although the idea of a social structure has already been presented previously [63], there was no indication of how each agent would build this social network representation. The only attempt made is in [6]. However, to the best of our knowledge

[3] For more details, see [34]

the proposed model has never been implemented or validated.

We have presented a method for agents to build a social network representation of their local environment. Using interaction information such as reputation information, agents can maintain their own representation of such environments. With this extended perception of the environments, agents can make more informed decisions. Results show, on average, an improvement of nearly 3% in the quality of their interactions, over a total of more than a million interactions. With this approach, we strive towards building an archetypal model for trust by combining the concepts of social networking and trust and reputation relationships.

Although the proposed method for identifying social relationships is specific to the particular scenario used for the evaluation, we should consider the general idea behind this. In a general market-like scenario, sellers who sell the same products, or provide the same kind of service, compete for the same market share, therefore they can be considered in competition, as well as agents needing a particular product or service, to operate in the market, need the "cooperation" of those agents selling that particular product or providing that service.

Our quantitative results and conclusions are based on simple observations: in Figures 4.3, 4.4, and 4.5 for instance, the results for the social agent clearly dominate those for the non-social agent under each scenario. For the other figures, more subtle conclusions were drawn. But in all, our results set the scene for a more systematic analysis using an in-depth statistical toolbox on the results for the various scenarios, and it indicates which parameters are interesting to analyse further.

We see this work as being influenced by but also contributing to further development of a theoretical framework for social structures in multi-agent systems. We have indicated in Section 3.1 how models of social structure are beginning to be considered in the MAS community, and our work gives some hints as to how they can be implemented and validated.

Chapter 5

An Abstract Framework for Trust in Coalition Formation

5.1 Motivation

The goal of coalition formation is typically to form robust, cohesive groups that can cooperate to the mutual benefit of all the coalition members. In open distributed systems, where there are many components that can enter and leave the system as they wish, the notion of *trust* becomes key when it comes to decisions about which coalitions to form, and when. When such systems are inhabited by agents that encompass some level of autonomy, each representing their own stakeholders with their own objectives, it not only becomes plausible that some agents are not trustable, but the consequences of joining a coalition of which some members cannot be trusted, or do not trust each other, become a key aspect in the decision of whether or not to join a group of agents.

With a relatively small number of exceptions mentioned in Chapter 2, existing models of coalition formation do not generally consider trust [17, 41]. In more general models [56, 38], individual agents use information about reputation and trust to rank agents according to their level of trustworthiness. Therefore, if an agent decides to form a coalition, it can select those agents it reckons to be trustworthy.

These models are inherently *local*: they lack a *global* view. They only consider the trust binding the agent starting the coalition and the agents receiving the request to join the coalition. In this chapter, we address this restriction. We propose an abstract framework through which autonomous, self-interested agents can form coalitions based on information relating to trust. In fact, we use *distrust* as the key social concept in our work. Luckily, in many societies, trust is usually regarded as the norm and distrust as the exception, so it seems reasonable to assume that a system is provided with information of agents that distrust each other based on previous experiences, rather than on reports of trust. Moreover, in several circumstances, it makes sense to assume that agents base their decision on which coalition to form on explicit information of distrust, rather than on information about trust. So, we focus on how distrust can be used as a mechanism for modelling and reasoning about the reliability of others, and, more importantly, about how to form coalitions that satisfy some stability criteria. We present several notions of mutually trusting coalitions and define different measures to aggregate the information presented in our distrust model.

Taking distrust as the basic entity in our model allows us to benefit in the sense of deriving our core definitions from an analogy with a popular and highly influential approach within *argumentation theory* [58]. Specifically, the distrust-based models that we introduce are inspired by the *abstract argumentation frameworks* proposed by Dung [26]. Other studies have used argumentation to reason about trust [62, 52], however our approach is different. These works are concerned with computing arguments with trust or look to construct arguments about the trust that one agent has in another. Our model does not use argumentation to reason about trust and does not use trust to build arguments; it uses Dung's abstract representation and interprets the relationships as distrust relationships.

In Dung's framework, an attack relation between arguments is the basic notion, which inspired us to model a distrust relation between agents. We show that several notions of stability and of extensions in the theory of Dung naturally carry over to a system where distrust, rather than attack, is at the core. We extend and refine some of these notions to our trust setting. In what follows, for each definition we provide, we will relate, if appropriate, to the notions of Dung's theory.

5.2 A Framework For Trust . . . Based On Distrust

In this Section, we introduce the basic models that we use throughout the remainder of this chapter. Our approach is inspired by the abstract argumentation frameworks of Dung [26]. Essentially, Dung was interested in trying to provide a framework that would make it possible to make sense of a domain of discourse on which there were potentially conflicting views. He considered the various conflicting views to be represented in *arguments*, with an *attack relation* between arguments defining which arguments were considered to be inconsistent with each other. This attack relation is used to determine which sets of arguments are acceptable in some specified way. Evaluating conflicting arguments gives rise to ”extensions” that denote minimally/maximally consistent sets of belief. These sets relate to the notion of conflict-freeness or compatibility between the arguments in each set. Notice that Dung's framework focuses on capturing the notion of two arguments being incompatible, or inconsistent with one-another, rather than supporting each other. In our work, we use similar graph like models, but our model is made up of agents, rather than arguments, and the binary relation (which is used in determining which coalitions are acceptable), is a *distrust* relation.

A *distrust* relation between agent i and agent j is intended as agent i having none or little trust in agent j. More precisely, when saying that agent i distrusts agent j we mean that, in the context at hand, agent i has insufficient confidence in agent j to share membership with j in a coalition.

When the agents share their evaluation about others, we can build a distrust network of agents linked by *distrust* relationships. Previous work about networks of agents use trust relationships [2]. However these networks can be transformed into distrust-based networks. Trust values are usually binary or they fall in a fixed range. When trust can assume values from a fixed range, one way to infer distrust relationships starting from trust relationships is to set a threshold below which the trust value is considered too low to trust the agent it is referring to.

The follow definitions characterize our formal model.

Definition 1. *An* Abstract Trust Framework *(*ATF*), S, is a pair:*

$$S = \langle Ag, \rightsquigarrow \rangle$$

where:

- *Ag is a finite, non-empty set of* agents*; and*
- $\rightsquigarrow\ \subseteq Ag \times Ag$ *is a binary* distrust *relation on Ag.*

When $i \rightsquigarrow j$ we say that agent i distrusts agent j. We assume $\rightsquigarrow$ to be irreflexive, i.e., no agent i distrusts itself. Whenever i does not distrust j, we write $i \not\rightsquigarrow j$. So, we assume $\forall i \in Ag$, $i \not\rightsquigarrow i$. We call an agent i fully trustworthy *if for all $j \in Ag$, we have $j \not\rightsquigarrow i$. Also, i is* trustworthy *if for some $j \neq i$, $j \not\rightsquigarrow i$ holds. Conversely, call i* fully trusting *if for no j, $i \rightsquigarrow j$. And i is* trusting *if for some $j \neq i$, $i \not\rightsquigarrow j$* [1] .

The assumption according to which an agent never distrusts itself might seem too strong in a coalition formation context. Coalitions are usually created to allow the member to achieve a goal that could not be achieved by each agent singularly. Achieving this goal might require carrying out tasks that a single agent is not equipped for. Therefore, it is possible that an agent could not trust itself to carry out a particular task. However, at this stage, we are abstracting from the particular task or goal that the coalition is trying to achieve. Hence, we are not taking into account any skills that agents must possess to achieve the goal.

Later, we will find it convenient to compare abstract trust frameworks, and for this we use the following definition.

Definition 2. *If $S_1 = \langle Ag_1, \rightsquigarrow_1 \rangle$ and $S_2 = \langle Ag_2, \rightsquigarrow_2 \rangle$ are two* ATF*s, we say that S_2 extends S_1, written $S_1 \sqsubseteq S_2$, if both $Ag_1 \subseteq Ag_2$ and $\rightsquigarrow_1\ \subseteq\ \rightsquigarrow_2$.*

Let us consider an example.

[1] This assumption differentiates our model from Dung's who allowed for self attacking arguments.

Example 1. *Consider the* ATF*s in Figure 5.1, Figure 5.2 and Figure 5.3. In our representation, the vertices represent agents and the arrows between them represent the distrust relation: if* $a \rightsquigarrow b$*, we represent this by an arrow from a to b. Now:*

- *In the* ATF*s* S_1 *and* S_2*, distrust takes the pattern of a cycle. In particular, all agents are trustworthy (but no agent is fully trustworthy) and all agents are trusting (but none is fully trusting).*
- *In* S_3*, agent a is not trusting and is not trustworthy.*
- ATF S_4 *represents a situation where agents may be physically located linearly, and no agent trusts any of its neighbours.*
- *In* S_5*, agents a and d are fully trustworthy and c and d are both fully trusting.*
- *The societies* S_6 *and* S_8 *are like* S_2*, but now there is an additional agent d who, in* S_8 *is the only fully trustworthy agent, while in* S_6 *he is the only fully trusting agent.*
- *The society* S_7 *is an extension of* S_6*. In this society, no agent is fully trustworthy neither fully trusting. All agents distrust and are distrusted by other agents.*

5.2.1 Coalitions with Trust

In what follows, when we refer to a "coalition" it should be understood that we mean nothing other than a subset C of Ag. When forming a coalition, there are several ways to measure how much distrust there is among them, or how trustable the coalition is with respect to the overall set of agents Ag.

Definition 3. *Given an* ATF $S = \langle Ag, \rightsquigarrow \rangle$*, a coalition* $C \subseteq Ag$ *is* distrust-free *if no member of C distrusts any other member of C. Note that the empty coalition and singleton coalitions* $\{i\}$ *are distrust-free: we call them trivial coalitions.*

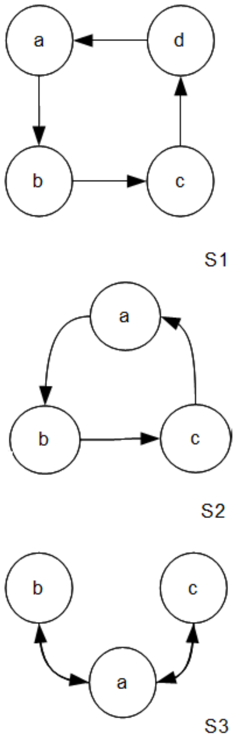

Figure 5.1: Three simple ATFs

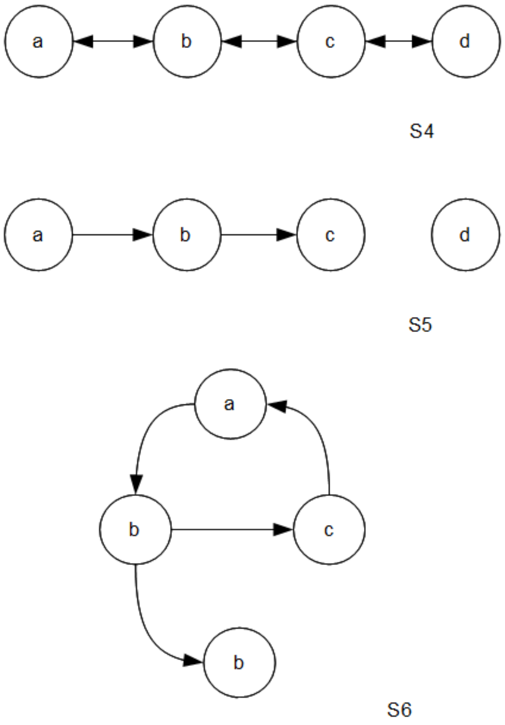

Figure 5.2: Three ATFs for four agents

Distrust freeness can be thought of as the most basic requirement for a *trusted* coalition of agents. It means that a set of agents has no internal distrust relationships between them. This idea is similar to Dung's definition of conflict-free sets of arguments.

Since we assume $\rightsquigarrow$ to be irreflexive, we know that for any $i \in Ag$, the coalition $\{i\}$ is distrust-free, as is the empty coalition. A distrust-free coalition for S_5 in Figure 5.2 is, for example, $\{a, c, d\}$, and, while $\{b, c\}$ is distrust-free in S_3, society S_2 has no distrust-free coalitions other than the trivial ones.

Consider ATF S_5 from Figure 5.2. The coalition $C_1 = \{c, d\}$ is distrust-free, but

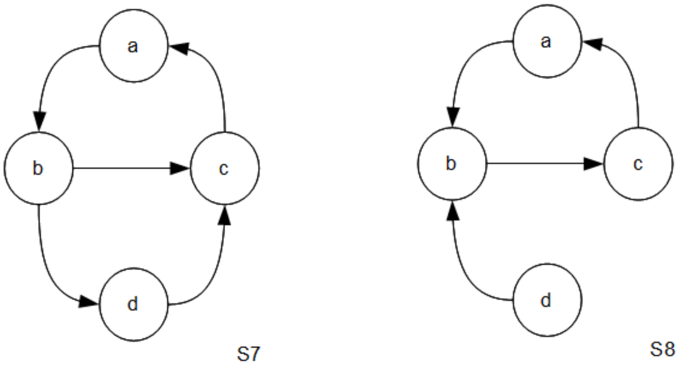

Figure 5.3: Two ATFs for four agents

still, they are not angelic: one of their members is being distrusted by some agent in Ag, and they do not have any justification to ignore that. Compare this to the coalition $C_2 = \{a,c,d\}$: any accusations about the trustworthiness of c by b can be neutralised by the fact that a does not trust b in the first place. So, as a collective, they have a defense against possible distrust against them.

With this in mind, we define the following concepts.

Definition 4. *Let* ATF $S = \langle Ag, \rightsquigarrow \rangle$ *be given.*

- *An agent* $i \in Ag$ *is called* trustable *with respect to a coalition* $C \subseteq Ag$ *iff* $\forall y \in Ag((y \rightsquigarrow i) \Rightarrow \exists x \in C(x \rightsquigarrow y))$.
- *A coalition* $C \subseteq Ag$ *is a* trusted extension *of* S *iff* C *is distrust-free and every agent* $i \in C$ *is trustable with respect to* C.
- *A coalition* $C \subseteq Ag$ *is a* maximal trusted extension *of* S *if* C *is a trusted extension, and no superset of* C *is one.*

Trusted extensions and maximal trusted extensions relate to the concept of admissible and preferred extensions in Dung's theory.

Consider the two ATFs S_1 and S_2 of Figure 5.1, where distrust takes the form of a cycle. In S_2, the only coalitions C with respect to which a is trustable are the coalitions that have b as a member: indeed, agent c distrusts a, but c in turn is distrusted by b. However, there is no trusted extension of S_2, which is easily seen as follows. Suppose a would be in a trusted extension S. Since c distrusts a, there needs to be an agent in S that distrusts c. The only agent that qualifies would be b, but b and a cannot be at the same time in a distrust-free coalition. Similarly, the only coalitions C with respect to which c is trustable are the coalitions that have a as a member: indeed, agent b distrusts c, but b in turn is distrusted by a. Again, suppose c would be in a trusted extension S. Since b distrusts c, there needs to be an agent in S that distrusts b. The only agent that qualifies would be a, but c distrusts a, therefore they cannot be at the same time in a distrust-free coalition. The same process applies to agent b.

Contrast this to the society S_1, where the distrust relation also forms a cycle, but where the two extensions are $\{a,c\}$ and $\{b,d\}$. It is easy to verify that those are also maximal trusted extensions. So the two cycles S_1 and S_2 demonstrate that an ATF can have several, or no maximal trusted extensions, respectively. Many studies have been carried out regarding results concerning the extensions yielded by odd and even length cycles in the abstract argumentation framework [55, 13]. Similar results could be found in our work.

In S_3 we have a society where agents have *mutual distrust* in each other. Here, a is trustable with respect to every coalition C that includes a as a member. Agent b is trustable with respect to $\{c\}$ and with respect to $\{b\}$ and all their supersets. Here, the maximal trusted extensions are $\{a\}$ and $\{b,c\}$.

Let us consider the societies of Figure 5.2. The maximal trusted extensions of S_4 are $\{a,c\}$ and $\{b,d\}$. In general, if we have n agents lined up as in S_4, the agents a_1 and a_n will be members of a maximal trusted extension, and such extensions will consists of an agent and all its neighbours of neighbours, and their neighbours of neighbours, etc. S_5 has a unique maximal trusted extension, $\{a,c,d\}$.

Finally, note that we have $S_2 \sqsubseteq S_6 \sqsubseteq S_7$. We have already seen that S_2 has no maximal extensions, and inspection shows that S_6 has neither. However, S_7 does have a maximal trusted extension: $\{a, d\}$. This shows that an extended society can gain extensions, as soon as *d* starts to distrust *c* in S_6, agents *a* and *d* together have a good story why they would work together, and not with anybody else.

The concept of a *trusted extension* represents a basic and important notion for agents who want to rationally decide who to form a coalition with, basing their decisions on trust. In particular: *a trusted extension is composed of agents that have a rational basis to trust each other*.

In Section 5.1, we underlined the fact that current coalition formation models based on trust [56, 38], only consider the trust relationship between the agent who asks another to join a coalition, and the agent asked. In our opinion, this approach neglects to consider a more *global* view of the coalition. It is important to take into account the trust relationships among all the agents in the coalition, to make sure that everyone is satisfied. Consider the following example.

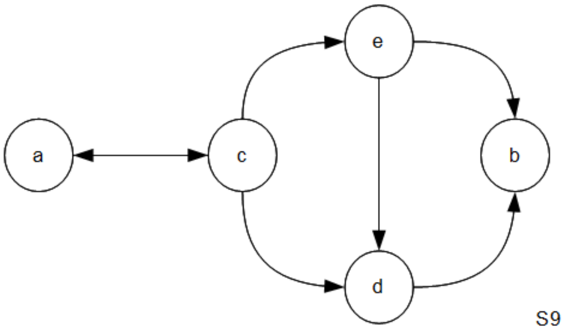

Figure 5.4: A more complex ATF

Example 2. *Consider the* ATF S_9 *in Figure 5.4. According to [56], the process of forming a coalition should be initiated by a single agent, who will ask other agents in the society whom he believes trustworthy to join him in a coalition. Consider agent* a *as the agent initiating the coalition formation process. Let us assume a is looking for at least three fellow agents to form a coalition with. According to the* ATF S_9*, a distrusts only c, therefore a can ask all the agents but c to form a* *coalition with it. Suppose a asks b and since b does not have a reason not to trust a, it accepts the invitation. We now have a temporary coalition C_1 formed by $\{a,b\}$. The process continues and a decides to ask e to join it. Again, e has no reason to distrust a so he accepts too. Our temporary coalition C_1 is formed by $\{a,b,e\}$. Continuing in this direction, a asks also d, who also has no reason not to trust a so it accepts the invitation. When a has selected the number of agents it needs, the process stops. Therefore the final coalition C_1 is formed by $\{a,b,e,d\}$. All agents in C_1 are trusted by a and they all trust a.*

In this example, it is easy to see that b is not trusted by d and both b and d are not trusted by e. Therefore if, for example, the agent starting the process of forming a coalition had been b, both e and d would have refused. Similarly, if e had been asked by d to form a coalition with it, it would have refused. Nevertheless, following this kind of approach, which uses trust as a factor in forming coalitions, the result would be a coalition where at least three of its members, e, b, and d, do not trust each other.

Coalition stability is a crucial problem. If agents do not trust the other components of the coalition, they could break away from the alliance, playing a negative role on the stability. Therefore, trust plays an important role for coalition stability. *Trusted extensions* provide a simple method for the agents to find *coalitions* where all the members are satisfied with their components.

5.2.2 Weak and Strong Trust

Maximal trusted extensions are a very interesting concept when considering the formation of trusting coalitions. As mentioned in Section 5.2.1, it is possible that a particular ATF has more than one maximal trusted extension. One could assume that all the agents in the maximal trusted extensions are equally trustworthy. For example, with regard to ATF S_4, in Figure 5.2 the maximal trusted extensions are $\{a,c\}$ and $\{b,d\}$. In this case, the agents appear respectively in only one maximal trusted extension. Now consider the following ATF S_{10}, as shown in Figure 5.5.

Here the maximal trusted extensions are $\{b,d,e\}$ and $\{b,f,c\}$. Suppose we are trying to determine the status of two agents, b and c. One way to address this is to

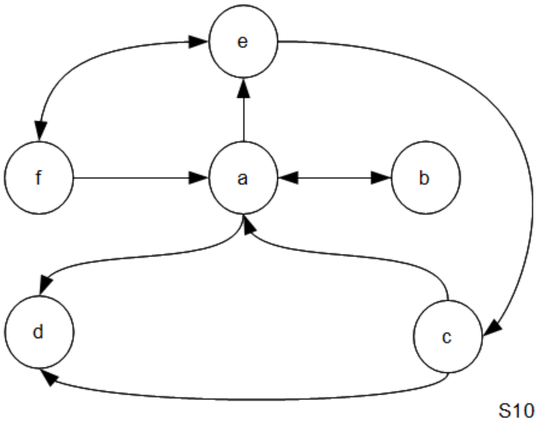

Figure 5.5: An ATF for weakly and strongly trusted agents

consider how many times *b* and *c* occur in a maximal trusted extension. We can see that *b* occurs in both maximal trusted extensions, while *c* occurs in just one of them. Therefore we can take this as evidence that *b* is somehow more “trustworthy” than *c*.

With this in mind, we define the following concepts.

Definition 5. *Let* ATF $S = \langle Ag, \rightsquigarrow \rangle$ *be given.*

- *An agent* $i \in Ag$ *is* Strongly Trusted *if it is a member of* every *maximal trusted extension.*
- *An agent* $i \in Ag$ *is* Weakly Trusted *if it is a member of at least* one *maximal trusted extension.*

Therefore, returning to the ATF in Figure 5.5, agents c, d, e, f are weakly trusted and agent *b* is strongly trusted (and hence also weakly trusted).

The notion of strongly and weakly trusted [2] can help agents decide in those

[2] The notion of strongly and weakly trust could be related to the Dung’s framework notion of credulous and sceptical acceptance [28]

situations where there are large maximal trusted extensions but not all the agents are required for forming a stable coalition .

5.3 Personal Extensions

In large societies, it is very unlikely that a single agent manages to interact with everyone in the society. For this reason, it has to rely on information given by others, about the reputation of the agents it does not know. Reputation can be defined as the opinion or view of someone about something [63]. This view can be mainly derived from an aggregation of opinions of members of the community about one of them. However, it is possible that the agent does not trust a particular agent and it wants to discard its opinion. Therefore, when it comes to forming a coalition, the agent wants to consider only its personal opinion and the opinion of the agent it trusts, while still keeping the coalition distrust-free.

For example, suppose that an agent wants to start a project and it needs to form a coalition to achieve its goals. It wants to form a coalition composed only of agents it trusts and who have no distrust relations among them.

To capture this intuition, we introduce two notions of *personal extensions*, which make it precise.

Maximal Personal Extensions: The first concept we define is the notion of the *Maximal Personal Extension*.

Definition 6. *Let the* ATF $S = \langle Ag, \leadsto \rangle$ *and* $a \in Ag$ *be given. Then* $C \subseteq Ag$ *is a Maximal Personal Extension generated by a, (notation:* $MPE(S,a)$*), if it is a* maximal trusted extension *of the* ATF $S' = \langle Ag', \leadsto' \rangle$*, where:*

- $Ag' = \{x \in Ag \mid x \not\leadsto a\}$*; and*
- $\leadsto' = (Ag' \times Ag') \cap \leadsto$.

That is, we remove all agents that distrust a from Ag, and restrict the distrust relation to that set.

Consider the ATF $S = \langle Ag, \leadsto \rangle$ in Figure 5.6. Suppose agent a wants to compute its personal extensions. Then, according to our definition, agents x and c will

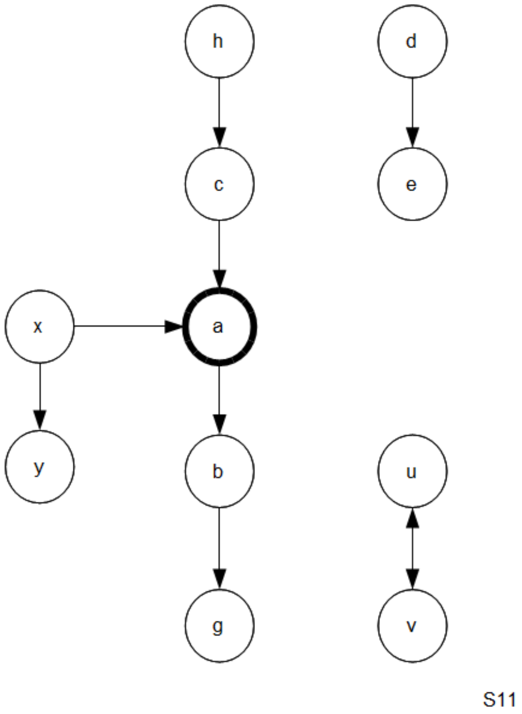

Figure 5.6: S_{11}, an ATF for Personal Extension

be discarded because they distrust a. Therefore, the *maximal trusted extensions* computed for this restricted set Ag' are $\{a,d,g,h,y,u\}$ and $\{a,d,g,h,y,v\}$. In general, we can say that an agent b enters a maximal personal extension as long as everybody that distrusts it is distrusted by someone which is accepted.

Unique Personal Extensions: Although the extensions obtained in this way are maximal (w.r.t. set inclusion), this definition allows for more than one maximal personal extension. Therefore, with regards to the example in Figure 5.6, agent a will have to choose between *two* personal extensions. However, without other information or additional criteria available to it, it would not be able to make a justified decision. Therefore we introduce our second notion of personal extension, the *unique personal extension*.

Given an ATF $S = \langle Ag, \rightsquigarrow \rangle$, and an agent $a \in Ag$, the unique personal extension $UPE(S,a)$ is required to have the following properties:

1. $a \in UPE(S,a)$
2. $UPE(S,a)$ is unique
3. $UPE(S,a)$ is distrust free
4. there is a minimal set $OUT \subseteq Ag$, with the following properties, for all agents $x, y \in Ag$:
   (a) $x \rightsquigarrow a \Rightarrow x \in OUT$
   (b) $(y \in UPE(S,a) \; \& \; y \rightsquigarrow x) \Rightarrow x \in OUT$
   (c) $y \in UPE(S,a) \Leftrightarrow \forall z(z \rightsquigarrow y \Rightarrow z \in OUT)$

So we make a coalition $UPE(S,a)$ for a in such a way that there is a minimal set OUT such that $UPE(S,a)$-members are only distrusted by OUT-members, and only those that distrust a or that are distrusted by a member of $UPE(S,a)$, are OUT. Loosely put: we add a to $UPE(S,a)$, and then we ensure that whoever distrusts or is distrusted by somebody in $UPE(S,a)$ is out, while $UPE(S,a)$ only accepts those agents as members that are at most distrusted by members of OUT.

We define $UPE(S,a)$ through an algorithm that generates it, from which the first three properties can be directly derived (see Figure 5.7). The algorithm works as follows. Given an ATF $S = \langle Ag, \rightsquigarrow \rangle$, we take an agent $a \in Ag$, for whom we want to compute the unique personal extension $UPE(S,a)$, with the idea that this extension is conceived as iteratively computing sets of agents *IN* (the agents accepted in the process) using sets OUT (the agents that are rejected), *CANDs* (agents not in OUT) and, finally, a set *PROMOTE*: those agents from *CANDs* that stand the test that they are not distrusted by any agent in *CANDs* and go to *IN*. Note that the properties $IN \subseteq CANDs$ and $CANDs = Ag \setminus OUT$ are *invariants* of the algorithm: they are both true at line 5, before entering the while-loop, and at line 10, at the end of the loop.

Initially, nobody is *IN* (line 2 of Figure 5.7), but we put *a*, the agent whose personal extension we are computing, in *PROMOTE* (line 3), while the agents which distrust a are definitely OUT. Then, as long as there are agents to be promoted, do the following: mark the agents to be promoted as *IN* , (line 7), and make those agents that are distrusted by any agent that is *IN* definitely OUT (line 8). Then remove the agents that are OUT from the set of *CANDs* (line 9) and *PROMOTE* those agents that are candidates but not yet in if they are not distrusted by any candidate in *CANDs* (line 10) [3].

Note that agents can be out for two reasons: first of all, they may distrust agent a (line 4), or they may themselves be distrusted by an agent that is in (line 8).

Theorem 1. *Let the* ATF $S = \langle Ag, \rightsquigarrow \rangle$ *and* $a \in Ag$ *be given.*

If we define $UPE(S,a)$ *as the set IN returned by the function generate-*$UPE(S,a)$*, then* $UPE(S,a)$ *satisfies the four requirements set out above.*

Proof. Membership of $a \in UPE(S,a)$ is because of lines 3 and 7, uniqueness is because the algorithm is deterministic and it terminates. The latter holds since every time $PROMOTE \neq \emptyset$ at line 10, the set *IN* strictly increases at line 7, and this can happen at most $|Ag|$ times. $UPE(S,a)$ is distrust-free because any candidate

[3] The algorithm introduced presents similarities with work carried out on labelling-based semantics for Dung's argumentation frameworks [68, 72]

```
1.  function $generate\text{-}UPE(\langle Ag, \leadsto \rangle, a)$ returns $UPE(S,a)$
2.      IN := $\emptyset$
3.      PROMOTE := $\{a\}$
4.      OUT := OUT $\cup \{b \in$CANDs $\mid b \leadsto a\}$
5.      CANDs := $Ag\setminus$ OUT
6.      while PROMOTE $\neq \emptyset$
7.          IN := IN $\cup$ PROMOTE
8.          OUT := OUT $\cup \{b \in$ CANDs $\mid \exists i \in$ IN $i \leadsto b\}$
9.          CANDs := CANDs $\setminus$ OUT
10.         PROMOTE := $\{c \in$ CANDs $\setminus$ IN $\mid \forall x \in$ CANDs $x \not\leadsto c\}$
11.     endwhile
12.     return IN
13. end-function
```

Figure 5.7: An algorithm for generating $UPE(S,a)$.

is definitely *OUT* if somebody in *IN* distrusts it (line 8). That there is a set *OUT* with the properties of item 4 is immediate from the algorithm: (a) is ensured on line 4, (b) on line 8, and (c) is true at line (5) and line (8). Since there are no other assignments to *OUT*, this set is a minimal set with the properties (a) - (c). □

The agents remaining in *IN* after this process form the agent *a*'s unique personal extension: $UPE(S,a)$.

Example 3. *Consider again the example shown in Figure 5.6. With respect to this example, we calculate UPE*(S,a)*.*

- *Initially, IN is the empty set, PROMOTE becomes* $\{a\}$ *and OUT becomes* $\{x\}$*, since x distrusts a;*

- *We then enter the while loop, during which in the first cycle, agent a enters IN, and OUT becomes* $\{x,b\}$*, since b is now distrusted by someone in IN. Everybody outside OUT is in CANDs and now the agents to be promoted are* $\{d,g,h,y\}$*: d and h are promoted since they are not distrusted by anybody,* *and g and y are promoted because the only agents that distrusted them (b and x, respectively), are now out.*

- *In the next cycle of the while-loop, the variable IN becomes* $\{a,d,g,h,y\}$ *and OUT is* $\{x,b,c,e\}$ *(the new members c and e are distrusted by the new IN-members h and d, respectively. The set CANDs now becomes* $\{a,d,g,h,y,u,v\}$*, and PROMOTE is now empty and the program terminates with IN* $= \{a,d,g,h,y\}$.

Note that the two agents u and v are candidates throughout the algorithm and never become members of IN or OUT.

In general, we have that $UPE(S,a)$ *is included in every personal extension generated by a set a. In fact, the two maximal personal extensions generated by a in the* ATF *of Figure 5.6 are* $UPE(S,a)\cup\{u\}$ *and* $UPE(S,a)\cup\{v\}$*. In* $UPE(S,a)$*, an agent b only enters if all agents distrusting b are already eliminated, while in the maximal personal extensions generated by a, we may allow some other agents, as long as everybody that distrusts them is distrusted by someone that is accepted. This yields to bigger coalitions, but, as shown, the result is not unique.*

All personal extensions, maximal and unique, are *distrust-free*. However, the distrust relationship is not symmetric, i.e., if *i* distrusts *j*, it is not necessarily the case that *j* distrusts *i*, and hence it can happen that one agent's personal extension is not *trusted* according with our definition 4. The agent preventing the extension from being trusted will be the agent who the personal extension belongs to. With respect to example in Figure 5.6, agent *a* is not *trustable*, as the agent distrusting it, agent *x* is not distrusted itself by any of the agents in the personal extension. However, for the purpose of the personal extension, this is not a problem because it represents the coalition that agent *a* would choose for itself. Clearly agent *a* considers itself trustworthy and agents who distrust it are not part of this coalition.

Another interesting property of the unique personal extension is illustrated by the following theorem.

Theorem 2. *Let* ATF $S = \langle Ag, \rightsquigarrow\rangle$*, be given. Then*
if $(\exists a \in Ag \mid \forall i \in Ag\ i \not\rightsquigarrow a)$ *then*
$(\forall j \in UPE(S,a), UPE(S,j) = UPE(S,a))$

Informally, Theorem 2 says that:

> *if the unique personal extension is computed by an agent who is distrust-free in the whole society, therefore it is fully trustworthy, then all the agents in that extension will have the same unique personal extension.*

Proof. Consider, for example, agent h's unique personal extension, from the ATF $S = \langle Ag, \leadsto \rangle$ in Figure 5.6. Following the algorithm in Figure 5.7, agent h's unique personal extension will be formed by $\{d,h,x,b\}$. If the UPE is computed by an agent who is distrust-free in the whole society, the unique personal extension is equivalent to computing the grounded extension in Dung's abstract argumentation framework. Every abstract argumentation framework has a unique grounded extension. Therefore all the agent $\in UPE(S,h)$ will have the same UPE. In fact, we can notice that $UPE(S,h) = UPE(S,x) = UPE(S,b) = UPE(S,d) = \{d,h,x,b\}$. □

It is important for the agent to know that, if it is not distrusted in the society, then the agents in its unique personal extension will share its vision of personal trustworthy coalition.

5.4 Aggregate Trust Measures

Abstract trust frameworks provide a social model of (dis)trust; they capture, at a relatively high level of abstraction, who (dis)trusts who in a society, and notions such as trusted extensions and personal extensions use these models to attempt to understand which coalitions are free of negative social views. An obvious question, however, is how the information presented in abstract trust frameworks can be *aggregated* to provide a single measure of how trustworthy (or otherwise) an individual within the society is. We now explore this issue. We present three aggregate measures of trust, which are given relative to an abstract trust framework $S = \langle Ag, \leadsto \rangle$ and an agent $i \in Ag$. All these trust values attempt to provide a principled way of measuring the overall trustworthiness of agent i, taking into account the information presented in S:

- *Distrust Degree*:

This value ignores the structure of an ATF, and simply looks at how many or how few agents in the society (dis)trust an agent.

- *Expected trustworthiness*:

  This value is the ratio of the number of maximal trusted extensions of which i is a member to the overall number of maximal trusted extensions in the system S.

- *Coalition expected trustworthiness*:

  This value attempts to measure the probability that an agent $i \in Ag$ would be trusted by an arbitrary coalition, picked from the overall set of possible coalitions in the system.

These latter two values are related to solution concepts such as the Banzhaf index, developed in the theory of cooperative games and voting power, and indeed they are inspired by these measures [32].

5.4.1 Distrust Degree

On the web, several successful approaches to credibility such as PageRank [18, 51] use methods derived from graph theory to model credibility, which utilize the connections of the resource for evaluation. Several graph theoretic models of credibility and text retrieval [63] rely on the consideration of the in-degree of the vertex, that is the sum of the incoming edges of that particular vertex in a directed graph. The degree of the incoming edges is used to extract importance and trustworthiness.

In our model, incoming edges are distrust relationships, therefore they represent a negative evaluation of a particular agent from the others in the society. Thus, measuring the in-degree of an agent in the society can give an indication of how reliable (or unreliable) that agent is considered overall.

Formally, we call this value the *distrust-degree* for an abstract trust framework $S = \langle Ag, \rightsquigarrow \rangle$ and an agent $i \in Ag$, denoted as $\delta_i(S)$, and it is defined:

$$\delta_i(S) = \frac{|\{x \mid x \in Ag \text{ and } x \rightsquigarrow i\}|}{|Ag|}.$$

This number provides us with a measure of the reliability of the agent in the whole society. The higher the number of agents that distrust it, the less reliable that agent is considered to be.

However, as we mentioned before, a maximal trusted extension or, in general, a coalition C, according to our approach, is a set of agents who trust each other. Therefore, these agents may not be interested in the evaluation of agents outside the coalition. They are more interested in a distrust degree relative to C. Hence, we define the following measure. The *coalition distrust-degree* for an abstract trust framework $S = \langle Ag, \leadsto \rangle$, a coalition C and an agent $i \in Ag$, denoted as $\delta_i^C(S)$, is defined as:

$$\delta_i^C(S) = \frac{|\{x \mid x \in C \; and \; x \leadsto i\}|}{|C|}.$$

The coalition distrust degree provides a measure for the agents in C to select agents outside the trusted coalition, who they believe to be more reliable among the agents in the society. Agents in C can rank the agents outside using the value of the coalition distrust degree. In this way, it is possible to obtain an ordered list of the agents who the coalition consider less unreliable. The smaller the value of the coalition distrust-degree, the more reliable the agent is considered.

5.4.2 Summary

So far, we have introduced four notions of extensions.

- **Trusted Extensions** The idea is to try and capture the concept of a group of agents willing to work together or cooperate because there are no internal conflicts among them;

- **Maximal Trusted Extensions** They are considered the largest trusted extensions achievable in a society with certain distrust relationships;

- **Maximal Personal Extensions** They try to frame the idea of trusted extension from a single agent point of view;

- **Unique Personal Extensions** They are derived from Personal Extensions and allow the agent to make decisions when more than one Personal Extension can be formed.

5.4.3 Expected Trustworthiness

As we noted above, the expected trustworthiness of an agent *i* in system *S* is the ratio of the number of maximal trusted extensions in *S* of which *i* is a member to the overall number of maximal trusted extensions in the system *S*. To put it another way, this value is the probability that agent *i* would appear in a maximal trusted extension, if we picked such an extension uniformly at random from the set of all maximal trusted extensions. Formally, letting $mte(S)$ denote the set of maximal trusted extensions in $S = \langle Ag, \rightsquigarrow \rangle$, the expected trustworthiness of agent $i \in Ag$ is denoted $\mu_i(S)$, defined as:

$$\mu_i(S) = \frac{|\{C \in mte(S) \mid i \in C\}|}{|mte(S)|}.$$

Clearly, if $\mu_i(S) = 1$ then *i* is strongly trusted, according to the terminology introduced above, and moreover *i* is weakly trusted iff $\mu_i(S) > 0$.

From existing results in the argumentation literature on computing extensions of abstract argument systems [29], we can also obtain the following:

Proposition 1. *Given an* ATF $S = \langle Ag, \rightsquigarrow \rangle$ *and an agent* $i \in Ag$*:*

1. *It is #P-hard to compute* $\mu_i(S)$.
2. *It is NP-hard to check whether* $\mu_i(S) > 0$.
3. *It is co-NP-hard to determine whether* $\mu_i(S) = 0$.

Proof. As mentioned in section 5.4, this measure is derived from the Banzhaf power index [32]. The Banzhaf index is derived by simply counting, for each member, the number of winning coalitions it can participate but which are not winning if it does not participate. The complexity of calculating the Banzhaf index has been proved in [46]. □

	$C \subseteq Ag \setminus \{i\}$	MPE in $C \cup \{i\}$	$\mu_i(C \cup \{i\})$
$i = a$	$\emptyset$	$\{a\}$	1
	$\{b\}$	$\{a\},\{b\}$	0.5
	$\{c\}$	$\{a,c\}$	1
	$\{d\}$	$\{a,d\}$	1
	$\{b,c\}$	$\{a,c\},\{b\}$	0.5
	$\{b,d\}$	$\{a,d\},\{b\}$	0.5
	$\{c,d\}$	$\{a,c\},\{a,d\}$	1
	$\{b,c,d\}$	$\{a,c\},\{b,d\}$	0.5
$i = b$	$\emptyset$	$\{b\}$	1
	$\{a\}$	$\{a\},\{b\}$	0.5
	$\{c\}$	$\{b\},\{c\}$	0.5
	$\{d\}$	$\{b,d\}$	1
	$\{a,c\}$	$\{a,c\},\{b\}$	0.5
	$\{a,d\}$	$\{a,d\},\{b\}$	0.5
	$\{c,d\}$	$\{b,d\},\{c\}$	0.5
	$\{b,c,d\}$	$\{a,c\},\{b,d\}$	0.5

Table 5.1: Table showing the values of μ_i to calculate the Coalition Expected Trustworthiness with regard to example S_4 in Figure 5.2

For example, with respect to S_4 in Figure 5.2, the maximal trusted extensions are $\{a,c\}$ and $\{b,d\}$, therefore the expected trustworthiness of all the agents in the maximal trusted extensions is 0.5, since each of them is present in only one of the two maximal trusted extensions.

However, in the example in S_{10} in Figure 5.5, the maximal trusted extensions are $\{b,d,e\}$ and $\{b,f,c\}$. Therefore the expected trustworthiness of agent b is 1, as it is a strongly trusted agent, while the expected trustworthiness of the other agents d, e, f, and c is 0.5.

5.4.4 Coalition Expected Trustworthiness

There is one obvious problem with the overall expected trustworthiness value, as we have introduced above. Suppose we have a society that is entirely trusting (i.e., the entire society is distrust free) apart from a single "rogue" agent, who distrusts everybody apart from himself, even though everybody trusts him. Then, according to our current definitions, there is no maximal trusted extension apart from the rogue agent. This is perhaps counter intuitive. To understand what the problem is, observe that when deriving the value $\mu_i(S)$, we are taking into account the views of *all* the agents in the society – which includes every rogue agent. It is this difficulty that we attempt to overcome in the following measure. To define this value, we need a little more notation. Where $R \subseteq X \times X$ is a binary relation on some set X and $C \subseteq X$, then we denote by $restr(R,C)$ the relation obtained from R by restricting it to C:

$$restr(R,C) = \{(s,s') \in R \mid \{s,s'\} \subseteq C\}.$$

Then, where $S = \langle Ag, \rightsquigarrow \rangle$ is an abstract trust framework, and $C \subseteq Ag$, we denote by $S \downarrow C$ the abstract trust framework obtained by restricting the distrust relation $\rightsquigarrow$ to C:

$$S \downarrow C = \langle C, restr(\rightsquigarrow, C) \rangle.$$

Given this, we can define the *coalition expected trustworthiness*, $\varepsilon_i(S)$, of an agent i in given an abstract trust framework $S = \langle Ag, \rightsquigarrow \rangle$ to be:

$$\varepsilon_i(S) = \frac{1}{2^{|Ag|-1}} \sum_{C \subseteq Ag \setminus \{i\}} \mu_i(S \downarrow C \cup \{i\}).$$

Thus, $\varepsilon_i(S)$ measures the expected value of μ_i for a coalition $C \cup \{i\}$ where $C \subseteq Ag \setminus \{i\}$ is picked uniformly at random from the set of all such possible coalitions. There are $2^{|Ag|-1}$ coalitions not containing i, hence the first term in the definition.

Consider example S_4 in Figure 5.2. In Table 5.1 we have shown a break down of all elements necessary to compute the Coalition Expected Trustworthiness for all the agents in S_4. Note that the value of $\mu_a(S_4)$ and $\mu_d(S_4)$ will be the same and the value of $\mu_b(S_4)$ and $\mu_c(S_4)$ will be the same as well. This is due to the particular shape of S_4. Therefore it is easy to notice that $\mu_a(S_4)$ (and $\mu_d(S_4)$) is equal to 0.75, while $\mu_b(S_4)$ (and $\mu_c(S_4)$) is equal to 0.625.

We can observe that the values of the expected trustworthiness and the coalition expected trustworthiness differ. The coalition expected trustworthiness value arguably gives us a clearer overall idea of what the trustworthiness of an agent would be with respect to the maximal trusted extensions that can potentially be formed, therefore it offers a better insight into trust issue related to the problem of forming coalitions.

5.5 Summary

In this chapter, we have addressed some of the limitations of existing trust-based coalition formation approaches. We have taken the notion of *distrust* to be our key social concept.

The main contribution of this part of the work is the definition of an abstract framework that overcomes the limitations mentioned in Section 5.1 allowing the agents to form distrust-free coalitions. We have also formulated several notions of mutually trusting coalitions. We have presented techniques for how the information presented in our distrust model can be aggregated to produce individual measures to evaluate the trustworthiness of the agent with respect to the whole society or to a particular coalition.

The ATF model is not utility based, so we are not considering stability from a utility-theoretic point of view, but there is nevertheless an interesting relationship between our notion of stability and that of cooperative game theory. The best-known notion of stability in cooperative game theory is the *core*: roughly, this solution concept considers a coalition to be stable if no subset of the coalition has any rational incentive to defect from the coalition, in the sense that they could earn more for themselves by defecting. In our view, a coalition is stable if no agent has any rational reason to distrust any of the members.

Overall, these notions and measures provide the agents with an effective tool to assess the reliability of the agent as an individual in a society and as part of a coalition.

Chapter 6

A Framework for Trust and Distrust (ATDF): An extension for ATF

6.1 Motivation

As explained in Chapter 5, agents in a society may not trust each other. These conflicting relationships are captured in the ATF model by the distrust relationships and are considered as a negative form of relationships. The concept of *trustable* agent has been introduced to reinstate some of the distrusted agents, that is those whose distrusters are in turn distrusted. This way, in a situation like *i* distrusts *j* and *j* distrusts *z*, it could be implicitly assumed that *i*, distrusting the agent that distrusts *z* is offering some kind of support, or indirect trust, to *z*, therefore this implicit relationship can be considered a positive one.

In the basic ATF, only distrust relationships are explicitly represented by the *distrust* relation, and positive relations are implicitly represented by the notion of indirect support. Therefore, this kind of trust and distrust are *dependent* concepts.

However, relations represented like this are not always a correct description of the trust relationships in a realistic situation.

Many researchers have explored similar problems relatively to argumentation. Some work is presented in [5]. However many argue that ”support” relationships are not needed in an argumentation framework.

However, in our context, a situation where *i* trusts *j* and a situation where *i* neither trusts *j* nor distrusts *j* cannot be considered equivalent situations. The absence of any kind of relation could simply mean that *i* has no opinion about *j*.

Hence, a richer framework is called for to formalize situations where two *independent* kinds of relations are available.

The next Section gives the formal definition of the framework extended with the explicit *trust* relationship.

6.2 A Framework for trust and distrust: A Formal Definition

In this Section, we present an extension of the AFT described in Section 5. As mentioned in Section 6.1, in an AFT, the *absence of a distrust relationship* between two agents does not necessarily imply the *presence of a trust relationship*. Therefore it seems natural to complement the distrust framework just introduced with trust relationships.

In this framework, as before, we make no assumptions about why agents share their evaluations of the other agents in the community.

A *trust* relation between agent *i* and agent *j* formalises the idea of an agent *i* having some trust in agent *j*. More precisely, when saying that agent *i* trusts agent *j* we mean that, in this context, agent *i* believes to have sufficient confidence in agent *j* to share membership with *j* in the same coalition. When the agents share their evaluation about trust and distrust toward others, a social network of agents linked by *distrust* and *trust* relationships can be built.

The following definitions characterize our formal model.

Definition 7. *An* Abstract Trust/Distrust Framework *(*ATDF*), S, is a triple:*

$$S = \langle Ag, \rightsquigarrow, \dashrightarrow \rangle$$

where:

- *Ag is a finite, non-empty set of* agents*; and*

- *$\rightsquigarrow \ \subseteq Ag \times Ag$ is a binary* distrust *relation on Ag, as described in Section 5.2.*

- *$\dashrightarrow \ \subseteq Ag \times Ag$ is a binary* trust *relation on Ag.*

- *$\forall i, j \in Ag$ if $i \rightsquigarrow j$ then $i \not\dashrightarrow j$;*

- *$\forall i \in Ag : \ i \dashrightarrow i$;*

When $i \dashrightarrow j$ we say that agent i trusts agent j. We assume $\dashrightarrow$ to be always reflexive, i.e., every agent always trusts himself. So, we assume, without drawing the link, that $\forall i \in Ag$, $i \dashrightarrow i$. Whenever i does not trust j, we write $i \not\dashrightarrow j$. Call an agent i fully trusty *if for all $j \in Ag$, we have $j \dashrightarrow i$. Also, i is* trusty *if for some $j \neq i$, $j \dashrightarrow i$ holds. Conversely, call i* fully trustful *if for all j, $i \dashrightarrow j$. And i is* trustful *if for some $j \neq i$, $i \dashrightarrow j$.*

Also in this extension, as well as in the ATF, the absence of *trust* relationships between two agents i and j does not imply the presence of a *distrust* relationship. Note that the fourth bullet of the definition states that an agent cannot trust and distrust another agent at the same time. However we do not impose any of the relationships to be mutual, therefore there is no restriction to prevent a situation where for two agents i,j, $i \dashrightarrow j$ and $j \rightsquigarrow i$. We also allow situations where $i \not\dashrightarrow j$ and $i \not\rightsquigarrow j$ at the same time.

Example 4. *Consider the* ATF*s in Figure 6.1 and Figure 6.2. We omit reflective trust link in figures. In our representation, the vertices stand for the agents and the straight-lined arrows between them represent the distrust relation and the dotted-lined arrows represent the trust relation. Now, in this* ATDF*, trust and distrust takes the pattern of a cycle. In particular, in S_{12}*

- *agents a and c are fully trusting but not trusty;*
- *agent b is fully trustworthy and trusty;*
- *agent a is trustful;*
- *but no agents are fully trusty or fully trustful.*

While in in S_{13}:

- *agents b and d are trusty, fully trustworthy and trusting;*
- *agents a and c are trustful, trustworthy and fully trusting;*

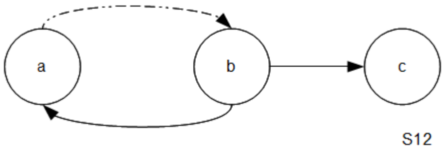

Figure 6.1: S12, an example of ATDF for three agents

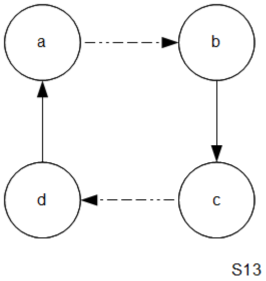

Figure 6.2: S13, an example of ATDF for four agents

6.3 Coalitions based on Trust

6.3.1 Trust and Transitivity

The relationship between *transitivity* and *trust* represents an important element to understand the intimate nature of the notion of trust. Since trust is considered the glue of social interactions, investigating its complex nature can help to realize a model to understand how it can be transferred to artificial societies [31].

Transitivity in trust is very often considered as a quite simple property, trivially inferrable from the classical transitivity defined in mathematics, logic, or grammar. The necessity of modelling trust in social networks is becoming more and more important, therefore research around this topic has been focused on the goal of inferring trust relationships between individuals that are not directly connected in the networks.

The problem of trust transitivity can be expressed as follows: if i trusts j, and j trusts z, what can be said about the trust relationship between i and z? Different and occasionally contrasting answers were given to this problem. The question is not only theoretically relevant; it is very relevant from a practical point of view [31]. If agents act in an open world, they will necessarily need to rely on third party information when interacting with agents they have not met before.

Many authors [24, 4] have questioned whether the transitive property can be applied to trust. Many other authors [42, 37, 31] analysed the problem and developed algorithms and solutions for inferring trust among agents not directly connected. These approaches, often, differ from each other in the way they compute trust values and propagate those values among the agents.

We adopt the point of view mentioned in [37]. Trust is not perfectly transitive in the mathematical sense, that is, if i highly trusts j, and j highly trusts z, it does not exactly follow that i will highly trust z. There is, however, a notion that trust can be passed between different individuals. When people ask a trusted friend for an opinion about a mechanic, they are taking the friend's opinion and use it to help form a preliminary opinion of the mechanic [37]. We use this idea in the concept of Trust Coalition, as explained in the next section.

6.3.2 T-Coalitions (Trust-Coalitions)

In this work, trust transitivity is not explicitly used to create new trust relationships among agents. However the notion of *T-Coalition* encompasses the idea of considering the opinion of friends of the agent.

The idea of the ATDF is to allow agents, who share their trust and distrust relationships, to form coalitions where all the members are satisfied with respect to the trustworthiness of the agents. We have already seen, in Chapter 5, different notions of mutually trusting coalitions, based on the principle of distrust freeness. As explained previously, distrust freeness can be thought of as the most basic requirement for a *trusted* coalition of agents. It means that a set of agents has no internal distrust relationships between them.

In this section, we provide a new definition of coalitions based on the trust relationships within the ATDF. Adding the trust relationships in our framework requires the addition of another principle, which will take into account these explicitly defined trust links between the agents.

The Trust-Coalitions (*T-Coalitions*) have to satisfy these two principles:

- Distrust-Freeness: A T-Coalition must be distrust-free with respect to the distrust relationship in the ATF;
- Trust-supported: If two agents are in the same T-coalition, they must be directly or indirectly linked by a trust relationship.

These principles are consistent with the idea that a coalition is formed by agents who do not have internal conflicts and recognise each other's some merits, therefore trust, in order to achieve a common goal or accomplish a task.

The *Distrust-Freeness* principle ensures that all the agents in the coalition have no reason to distrust each other, while the *Trust-Supported* principle ensures that all the agents are linked directly or indirectly by trust relationships.

We now present the formal definition of coalitions based on *trust relationships*.

Definition 8. *Given an* ATDF $S = \langle Ag, \rightsquigarrow, \dashrightarrow \rangle$*, a coalition* $C \subseteq Ag$ *is a* T-Coalition (Trust-Coalition) if:

1. C is *distrust-free w.r.t. the* $\rightsquigarrow$ *relationship;*

2. $\forall i$ *and* $\forall j \in C, \exists l_1, l_2, \ldots l_n \in C \mid i \dashrightarrow l_1$ $(or\ l_1 \dashrightarrow i)$, $\ldots$, $l_{n-1} \dashrightarrow l_n$ $(or\ l_n \dashrightarrow l_{n-1})$, $l_n \dashrightarrow j$ $(or\ j \dashrightarrow l_n)$*;*

3. *C is maximal* $w.r.t.$ $\subseteq$ *among the sets satisfying 1 and 2.*

The first requirement, (1), is in agreement with the *distrust-freeness* principle and the second one, (2), ensures that the *trust-supported* principle is satisfied. The third requirement, (3), simply ensures that the biggest coalition satisfying (1) and (2) is created. This so formed *T-coalition* can be subsequently "filtered" to satisfy any additional requirements, if needed.
Note that the singleton coalitions $\{i\}$ can be *T-Coalitions* as there is always an implicit *trust* relationship from and to the agent itself: we call them trivial T-Coalitions.

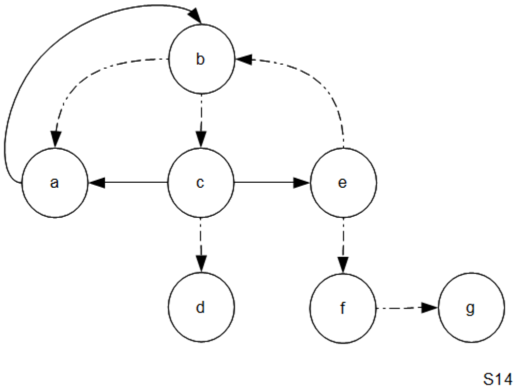

Figure 6.3: An example of T-Coalitions for an ATDFs

In $S14$, shown in Figure 6.3, the *T-Coalitions*, apart from the trivial ones, are $\{b, c, d\}$ and $\{b, e, f, g\}$.

We can see that b trusts a and c, and c trusts d. So, ideally b would be willing to work with a in the same coalition. However, a distrust b and c distrust a, therefore, adding a to the T-Coalition formed by $\{b,c,d\}$ would go against the distrust-freeness principle. The same can be said for the T-Coalition formed by $\{b,e,f,g\}$. If the distrust link between c and e was to be removed then the T-Coalition could be much larger, allowing also c and d to be in it; creating, this way, only one big large T-Coalition containing all the agents but a. The size of a coalition can be a very important factor in certain contexts, nevertheless, preserving the distrust-freeness is the most important requirement when considering coalition formation based on trust.

6.4 Trust-induced Inference Rules

We have seen in the previous subsection that the transitive property of the trust relationships is somehow encompassed by the use of the *T-Coalitions* which link the agents together in a coalition if they are tied by trust relationships, even if not directly.

Trust transitivity is normally used to infer an indirect trust relationship. For example, if i trusts j and j trusts z than we could say, in some cases, that it is reasonable to assume that i would trust z. The intuition is that, under normal circumstances, we trust whoever our friends trust.

The same idea could be applied on a more general scale. As mentioned at the beginning of this chapter, in Section 6.3.1, if we are friends with someone, we tend to trust his point of view towards his friends. Hence, it seems reasonable to assume that we tend not to trust whoever our friends distrust.

Therefore, we present a rule to infer *distrust* relationships from trusted agents.

Definition 9. *Given an* ATDF $S = \langle Ag, \rightsquigarrow, \dashrightarrow \rangle$*:*

$\forall i,j,z \in Ag$, *if* $i \dashrightarrow j$ *and* $j \rightsquigarrow z$ *and* $i \not\dashrightarrow z$ *then* $i \rightsquigarrow z$*;*

Note that, as mentioned in Definition 7, an agent cannot trust and distrust another agent at the same given time, and a direct relationships is considered "stronger" *than an inferred one. Therefore the* distrust *link is never inferred if there is already a* trust *relationship linking the two agents involved in the inference process. For the same reason, since it is implicitly assumed that an agent always trusts himself, the* distrust *relationship is not inferred if it would create a self-distrust link, creating a* *situation like $i \leadsto i$.*

The distrust relationship between the three agents is inferred only in the case there is a trust relation linking the first two agents, that it the reason why we call it *trust-induced*.

The inference process continues until there are no more *distrust* relations to be inferred. Once the process has stopped, the *maximal trusted extensions*, *personal extensions* and the *aggregate measures* can be re-computed.

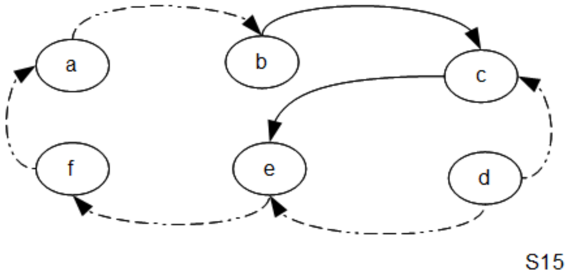

Figure 6.4: An example for the application of the inference rule. Step 0

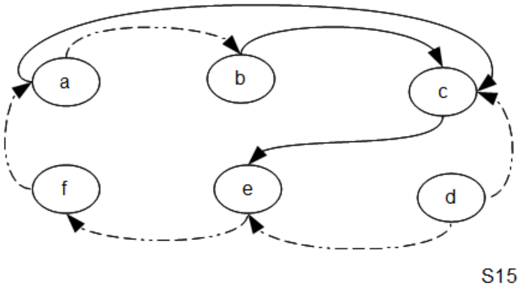

Figure 6.5: An example for the application of the inference rule. Step 1

Applying the *trust-induced inference rule* can add more information about the relationships among the agents and make the *distrust relations* clearly visible to the whole society, which can benefit from this additional knowledge.

In Figure 6.4 we can see an ATDF of six agents. We will use this society to show how the inference rule is applied. We call this *Step* 0, as it represents the initial status of the society. The *Maximal Trusted Extension* for this ATDF is $\{a,b,d,e,f\}$. Table 6.1 shows the aggregate measure values for each agent in the society for *Step* 0 and *Step* 3, the final one. In Figure 6.5, after applying the inference rule, we can notice that a distrust edge has been added between a and c, inferred from the fact that a trusts b and b distrusts c. Note that since d trusts c and c distrusts e, the inference rule would generate the creation of a distrust link between agent d and e. However d trusts e, therefore the inference rule is not applied in this case, as specified in Definition 9.

Now that a new distrust connection appeared, the inference rule can be applied again. Agent a now distrusts c, therefore, since agent f trusts a, a new distrust link can be created between agent f and c, as shown in Figure 6.6. Iterating the process, another distrust edge can be created between e and c. Note that, also in this case, the distrust link between d and c is not created because there exists already a direct trust relationships from d to c.

Notice that in the society at *Step* 3, the *Maximal Trusted Extension* remain unchanged, but the values of the aggregate measure changes for agent c and e. This suggests that these new *distrust* connections have provided the agents with additional information about the society that they can use to make their decisions.

The next chapter will describe the application of the framework to a practical scenario, and will show how the trusted extensions, the personal extensions and the aggregate measures change with the application of the inference rule.

6.5 Trends in Maximal Trusted Extensions

The addition of new *distrust* connections in a society, intuitively, suggests that the size of *distrust-free* groups of agents would possibly shrink but never increase. We have seen that, in the case we applied the inference rule, the number of agents in the society has not changed. Therefore, intuition tells us that, if we keep the size

i	Step 0		Step 3	
	$\mu_i(S)$	$\varepsilon_i(S)$	$\mu_i(S)$	$\varepsilon_i(S)$
a	1	1	1	1
b	1	1	1	1
c	0	0.5	0	0.1875
d	1	1	1	1
e	1	0.75	1	0.9375
f	1	1	1	1

Table 6.1: Table showing the values of μ_i to calculate the Coalition Expected Trustworthiness with regard to the example in Figure 6.4 for Step 0 and Figure 6.7 for Step 3

of the society unchanged, increasing the number of distrust links could only create more conflicts.

We are interested in verifying if it is possible to identify a trend in the way the size of the *Maximal Trusted Extensions* evolves with the increase of *distrust relationships*.

The intuition is that when the degree of *"connectivity"*[1] is 0%, there are no distrust relationships, therefore the *Maximal Trusted Extension* will be only one, of the size of the society itself. Conversely, when the *"connectivity"* degree is 100%, there are $n \times (n-1)$ distrust relationships, where n is the size of the society. In this graph, each agent distrusts, and is distrusted by, all the other agents, therefore we will have *n Maximal Trusted Extensions* of size 1, that is, every agent will be alone in one MTE.

Given these two extremes, we are interested in checking if there exists any sort of observable trend in between these two points.

It is easy to show that the trend in the changes of the average size of the MTEs and the number of MTEs are not monotonic. We use the ATF in Figure 6.8 and Figure 6.9 as an example to show that, effectively, the behaviour of the average size of the MTEs and the number of MTEs are not monotonic.

[1] We call *"connectivity"* degree the sum of the indegree and outdegree of the verteces of the graph

Step	*Mtes*	Number of *Mtes*	Average Size of *Mtes*
1	$\{a,b,c\}$	1	3
2	$\{a,c\}$	1	2
3	$\{b,c\}$	1	2
4	$\{\}$	1	0
5	$\{c\}$	1	1
6	$\{b\}$, $\{c\}$	2	1
7	$\{a\}$, $\{b\}$, $\{c\}$	3	1

Table 6.2: Table showing changes in *Mte*s for ATF in Figure 6.8 and Figure 6.9

Figure 6.8 and Figure 6.9 show what happens in the society as the distrust relationships increase. Table 6.2 shows the MTEs, and their number and average size at each step of the process of adding more distrust relationships.

We can notice that at step 1 with the connectivity equal to 0, we have only one MTE formed by all the agents. At step 2, adding one distrust relationship causes the MTE to shrink losing the one distrusted agent. At step 4, the distrust relationships take the shape of a circle, causing the MTE to disappear completely. This is the same scenario we have observed in the Figure 6.13 for series of societies of size 11 and 17. However, the addition of even one more distrust relationship can change the situation. In Step 5 of Figure 6.9 adding just one distrust relationship allows the formation of the MTE again. This time the size of the MTE has reached its minimum. In fact, we can notice that in the next steps, the number of MTEs increases but the average size remains unchanged.

We have carried out a number of experiments to verify if there is a trend, although not monotonic, in the way the size of the *Maximal Trusted Extensions* evolves with the increase of *distrust relationships*.

These tests do not consider the different topology configurations of the distrust relationships that the ATF can assume. Testing what happens in all possible topology configurations would be unfeasible. The number of topology configurations corresponds to the powerset of the set of all possible edges in the graph representing the ATF. We mentioned already that the size of this set is $n \times (n-1)$, therefore the size of this *powerset* would be $2^{n\times(n-1)}$. The complexity of these tests is clearly exponential.

Therefore the experiments we have carried out are not to be taken as a complete analysis of the behaviour of the *Maximal Trusted Extensions*.

We have tested the behaviour of the MTE on a number of societies of different sizes. We used different *probability* to increase the number of *distrust* links in the society. We gradually increased the probability that two random agents in the society are connected by a distrust relationship [2] from 0%, which gives a completely disconnected graph, where no agent distrusts any other agent, to 100%, which gives a completely connected graph, where all agents distrust each other.

We have measured what happens to the number of MTE and to the average size of the MTE, in societies of size starting from just one agent up to 20. The *probability of connectivity* is incremented 5% at each time step measured.

In Figure 6.10 the variation of the number of *Maximal Trusted Extensions* with the increase of the probability of connectivity is represented. We can observe that the chart does not show a smooth surface, which suggests that there is no monotonically increasing trend in the behaviour of the MTE. To better illustrate the point, we can observe Figure 6.11. This chart shows the behaviour of the MTEs for selected society sizes. We can see that all the series show a fluctuating trend until they stabilize to the maximum extreme.

Figure 6.12 shows how the average *size* of the *Maximal Trusted Extensions* varies with the increase of the probability of connectivity. In this chart we can observe that the average size of the MTEs shows a more monotonically-like decreasing behaviour.

Figure 6.13 shows selected series for certain society sizes, the same as in Figure 6.11. We can also observe that for the series of society size of 11 and 17 agents, the series has value 0 when the probability of connectivity has, respectively, values of 30% and 20%. This clearly means that the particular topology configuration of the distrusts relationships that the ATF assumes in that timestep creates conflicts such that no *trusted extensions* are possible.

[2] In what follows, we call this *probability of connectivity*

In conclusion, we can say our experiments showed that while the number of *Maximal Trusted Extensions* shows a high variation throughout the process of increasing the probability of connectivity, the average size of *Maximal Trusted Extensions* follows a smoother decreasing trend, only occasionally interrupted.

This indicates that a society with a high degree of distrust has less probability of being able to form a large trusted coalition.

6.6 Summary

In this chapter, we have presented an extension of the AFT described in Section 5. As mentioned in Section 6.1, in the AFT, the *absence of a distrust relationship* between two agents does not necessarily imply the *presence of a trust relationship*. Therefore it is natural to integrate trust relationships along with distrust relationships in the ATF framework.

The contribution of this extension is to allow the agents to explicitly specify which agent they trust and which one they distrust. The notion of *T-Coalitions* allows the agents to form distrust-free coalitions with members who are linked by trust relations.

Research on coalition stability is a crucial coalition problem. Stability is the motivation of an agent's refusal to break from the original coalition and form a new one. There are two factors determining an agent's trust in a coalition, one is that the agent must possess necessary competence for a particular task, the other is that all the agents involved in a coalition must behave honestly and diligently [56].

In this work, we have abstracted from the competence skills required to accomplish a particular task, but we concentrated on the belief that the agent needs to have that its fellow coalition members will perform the task assigned without defecting.

Therefore, these coalitions can be considered stable in the sense that no agent has any incentive to break from the coalition to join another one which it deems to be more reliable.

Moreover, the *trust-induced inference rule* defined in Section 6.4 can be compared to the use of reputation by agents when they do not have information from direct interactions. In a large society it is not possible to interact with everyone and form an opinion about all the agents. Therefore agents often rely on third party information about the agents they do not know personally. Applying the *trust-induced inference rule* can not only add more information about the relationships among the agents but can also make them clearly evident to the whole society, who may benefit from this additional knowledge.

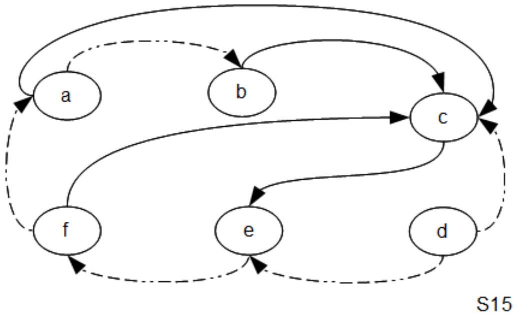

Figure 6.6: An example for the application of the inference rule. Step 2

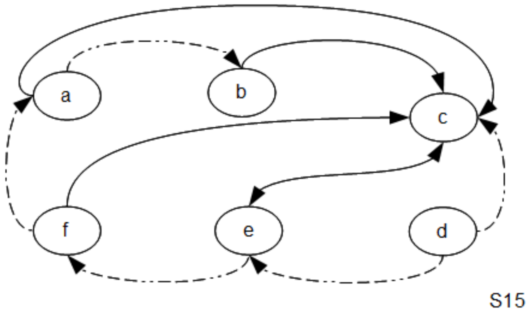

Figure 6.7: An example for the application of the inference rule. Step 3

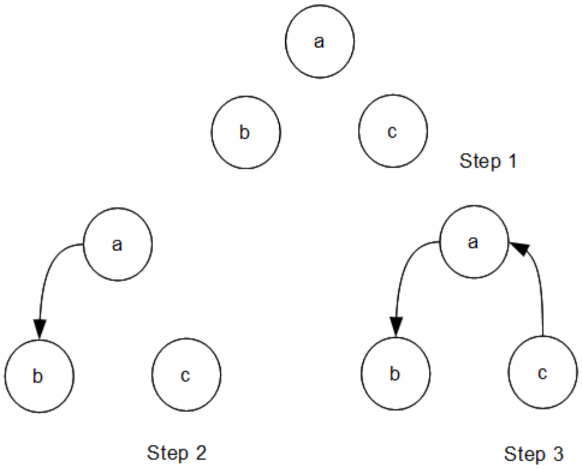

Figure 6.8: An example of changes in *Maximal Trusted Extensions* in a society of three agents with increasing number of *distrust relationships*. Step 1 to 3

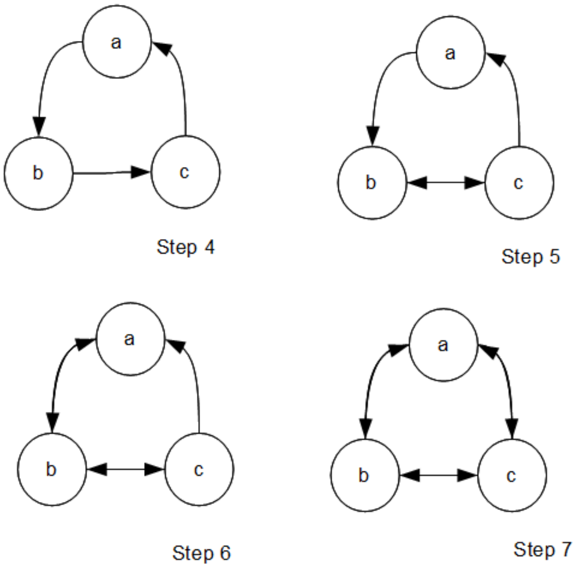

Figure 6.9: An example of changes in *Maximal Trusted Extensions* in a society of three agents with increasing number of *distrust relationships*. Step 4 to 7

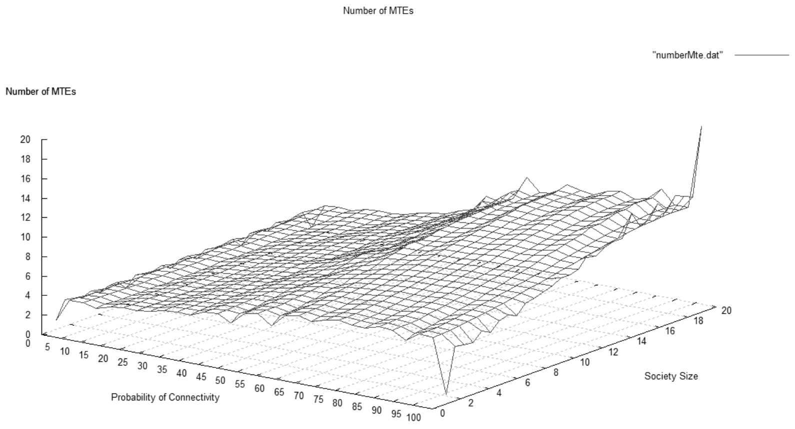

Figure 6.10: Chart showing how the number of Maximal Trusted Extensions changes with the increase of Distrust Relationships, for different society sizes

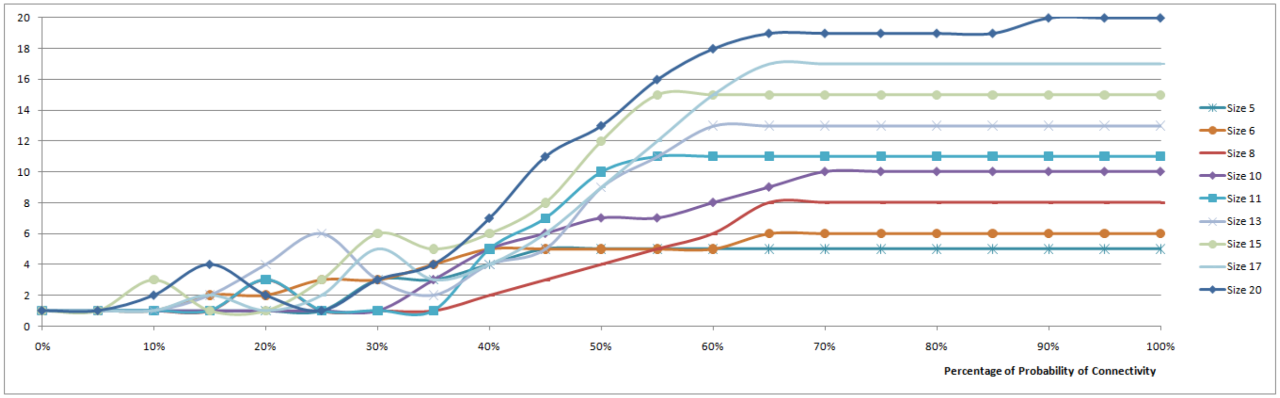

Figure 6.11: Chart showing how the number of Maximal Trusted Extensions changes with the increase of Distrust Relationships, for selected society sizes

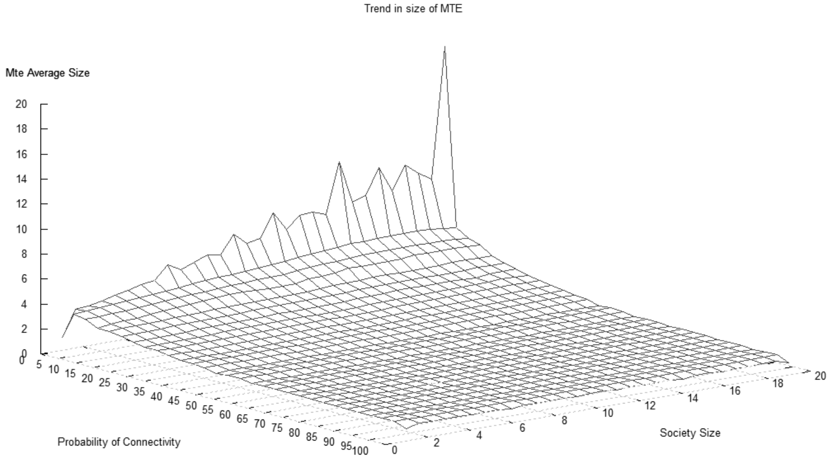

Figure 6.12: Chart showing how the average size of Maximal Trusted Extensions changes with the increase of Distrust Relationships, for different society sizes

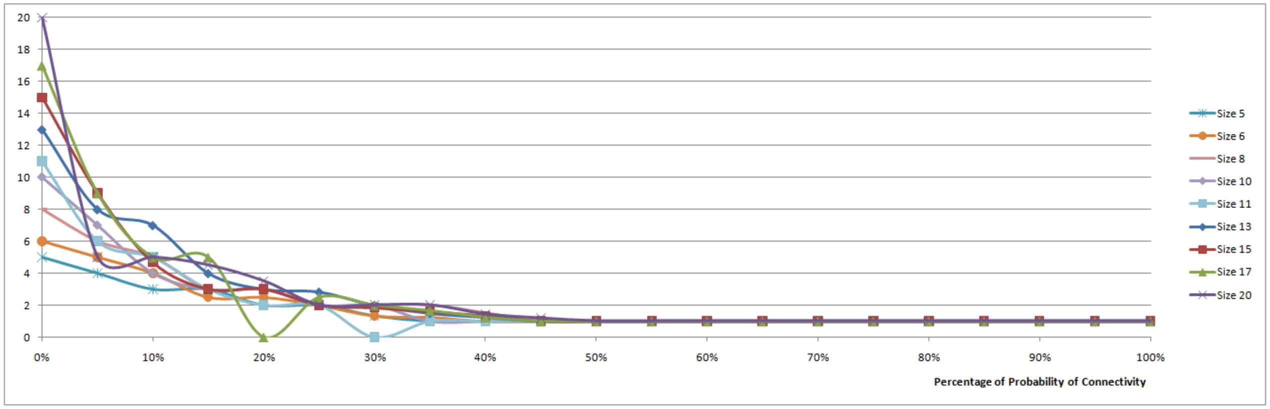

Figure 6.13: Chart showing how the average size of the Maximal Trusted Extensions changes with the increase of Distrust Relationships, for selected society sizes

Chapter 7

A Formal Analysis of Trust and Distrust Relationships in Shakespeare's Othello

In this chapter we use the famous tragedy by Shakespeare, Othello, to give an actual illustration and analysis of the models proposed.

We have implemented the ATF framework and the functionality necessary to extract the *Maximal Trusted Extensions*, *Personal Extensions* and compute all the *aggregate measures*.

The play [1] revolves around four central characters: Othello, a Moorish general in the Venetian army; his wife Desdemona; his lieutenant, Cassio; and his trusted ensign Iago. Othello is a general in the Venetian Republic military, who is in command of the army fighting the Turks attacking Cyprus.

The play opens with Roderigo, a rich and dissolute gentleman, complaining to Iago, a highly-ranked soldier, that Iago has not told him about the secret marriage between Desdemona, the daughter of a Senator named Brabantio, and Othello. Roderigo is upset by this development because he loves Desdemona and had previously asked her father for her hand in marriage. Iago is upset with Othello for promoting Cassio, a younger man, above him, and tells Roderigo that he plans to use Othello for his own advantage. Iago tries in many ways to have Cassio stripped of his rank but he also wants to ruin Othello for having preferred Cassio over him. Therefore Iago persuades Othello to be suspicious of Cassio and Desdemona. He achieves this using a precious handkerchief that was Othello's first gift to Desdemona and giving it to Cassio, with the help of his wife Emilia. The defense from Iago towards Cassio, although not sincere, and his deliberate reticence are the central part of Iago's work of persuasion which leads to Othello killing Desdemona in a fit of rage. In the epilogue, Emilia reveals that Desdemona's betrayal was invented by Iago, who immediately kills her. Othello, feeling guilty for killing the innocent Desdemona, kills himself. Iago is arrested and Cassio takes Othello's place as general.

The play is divided in five acts. We take this division as the natural breaking points for modelling the scenarios with our AFT and AFDT. We design an AFT and AFDT for each act, using the main characters and the $\rightsquigarrow$ and $\dashrightarrow$ relationships, representing the distrust (or more in general dislike) and trust relationships among the characters.

For each of the main characters we give the aggregate trust measures for each act and for Othello and Iago, the primary actors, we also present the UPE. Moreover we analyse how the ATDF evolves with the application of the inference rule, and show how these changes are reflected in the aggregate trust measures.

In our ATFs representing the play Othello, we only consider the main characters: Othello, Desdemona, Iago, Cassio, Roderigo and, for the first act, Brabantio. Therefore our context is limited to the relationships among these figures.

Normally, in multi-agent systems scenarios, a coalition is a set of agents who may or may not work together to achieve a common goal or to earn higher utility [70]. In our example, the Othello play, we abstract from the goal of the coalition and we consider the proposed trusted set of agents simply as sets whose components trust, or have no reason to distrust each other. In the next paragraphs we provide a more extensive description of each act and then we provide a detailed analysis of the society.

7.1 ACT 1 Synopsis

Shakespeare's famous play of love turned bad by unfounded jealousy begins in Venice with Iago, a soldier under Othello's command arguing with Roderigo, a wealthy Venetian. Roderigo has paid Iago a considerable sum of money to spy on Othello for him, since he wishes to take Othello's girlfriend, Desdemona as his own.

Roderigo fears that Iago has not been telling him enough about Desdemona and that this proves Iago's real loyalty is to Othello rather than him. Iago explains his hatred of Othello for choosing Cassio as his lieutenant and not him as he expected. To regain Roderigo's trust, Iago informs Brabantio, Desdemona's father, of her relationship with Othello the "Moor", which enrages Brabantio into sending parties out at night to apprehend Othello for what in Brabantio's eyes must be an abuse of his daughter by Othello. Iago lies that Roderigo, and not himself, was responsible for angering Brabantio against Othello, Iago telling Othello that he should watch out for Brabantio's men who are looking for him. Othello decides not to hide, since he believes his good name will stand him in good stead.

We learn that Othello has married Desdemona. Brabantio and Roderigo arrive, Brabantio accusing Othello of using magic on his daughter. Othello stops a fight before it can happen but he is called away to discuss a crisis in Cyprus, much to the anger of Brabantio who wants justice for what he believes Othello has done to his fair Desdemona. The Duke is in council with several senators discussing their enemy, the Turks (Turkish people). Brabantio complains to the Duke that Othello bewitched his daughter and had intimate relations with her. Desdemona is brought in to settle the matter, Othello meanwhile explains how he and Desdemona fell in love. Desdemona confirms this and the Duke advises Brabantio that he would be better off accepting the marriage than complaining and changing nothing.

The Duke orders Othello to Cyprus to fight the Turks, with Desdemona to follow, accompanied by the trusted Iago. Roderigo despairs that his quest for Desdemona is over now that she is married, but Iago tells him not to give up and earn money instead; soon Desdemona will bore of Othello.

Alone, Iago reveals his intention to continue using Roderigo for money and his hatred of Othello (Othello picked Cassio and not Iago as his lieutenant). Iago explains that his plan to get revenge on Othello is to suggest to him that Cassio is sleeping with Desdemona.

7.2 ACT 1 Analysis

In Figure 7.1 and 7.2 we observe that the only distrust-free characters, at the end of the first act, are Desdemona and Roderigo.

The *maximal trusted extension* is formed by

$$\{Desdemona, Roderigo, Cassio, Brabantio\}.$$

Note that Othello, although a noble person at the beginning of our story, is hated and therefore distrusted by Iago, who feels betrayed by him, by Roderigo who feels Desdemona has been stolen from him by Othello, and by Brabantio, Desdemona's father, who does not approve of the couple's marriage.

	Maximal Trusted Extensions	T-Coalitions
Act 1	$\{Desdemona, Roderigo, Cassio,$ $Brabantio\}$	$\{Desdemona, Brabantion\}$ $\{Othello, Desdemona, Cassio\}$
Act 2	$\{Desdemona, Roderigo, Iago\}$	$\{Roderigo, Iago\}$ $\{Othello, Desdemona, Cassio\}$
Act 3	$\{Desdemona, Iago\}$	$\{Desdemona, Cassio\}$
Act 4	$\{Desdemona, Iago\}$	$\{Desdemona, Cassio\}$
Act 5	$\{Desdemona, Roderigo\}$	$\{Othello, Desdemona\}$ $\{Cassio, Desdemona\}$

Table 7.1: Table showing the Maximal Trusted Extensions and the T-Coalitions in the Othello's society S for each Act

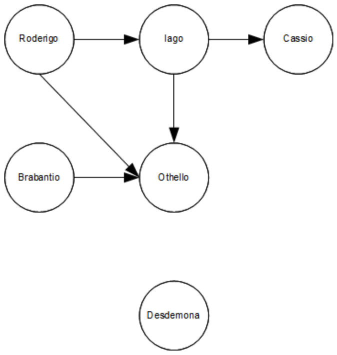

Figure 7.1: ATF for Act 1

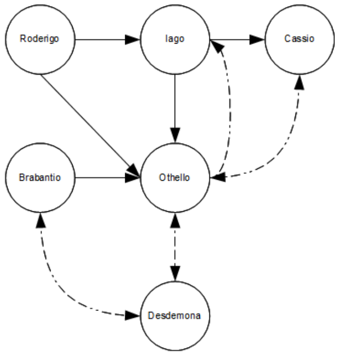

Figure 7.2: ATDF for Act 1

Therefore Othello is highly distrusted in the first act, as shown by his *coalition expected trustworthiness* value, 0.125, which is the lowest among all characters in this act. However, his beloved Desdemona and his loyal lieutenant Cassio trust him and, in fact they are part of Othello's *unique personal extension*. Desdemona, being distrust free is part of Iago's *unique personal extension*, together with Brabantio, who is also distrust free. The *T-Coalitions* in this first act are two:

$$\{Desdemona, Brabantio\} \text{ and } \{Othello, Desdemona, Cassio\}.$$

Note that Othello's *unique personal extension* corresponds exactly to the *T-Coalition* he belongs to. Although Iago's *coalition expected trustworthiness* value, 0.5, is higher than Othello's, we note that Iago does not belong to any *T-Coalition*. Although Othello trusts Iago, Iago himself despises Othello. This conflict prevents them from being together in a T-Coalition.

In Table 7.4, Table 7.5 and Table 7.6 we can find the values for the respective items for the society after applying the inference rule.

For the first act, it is not possible to infer any new *distrust* relationship. For instance, Desdemona trusts her father Brabantio and Brabantio distrusts Othello. The distrust relationship between Brabantio and Othello would cause Desdemona to adopt her father's point of view and distrust Othello, if we were to apply the

	Othello's UPE	Iago's UPE
Act 1	$\{Othello, Desdemona, Cassio\}$	$\{Iago, Desdemona, Brabantio\}$
Act 2	$\{Othello, Desdemona, Cassio\}$	$\{Iago, Desdemona, Roderigo\}$
Act 3	$\{Othello\}$	$\{Iago, Desdemona\}$
Act 4	$\{Othello\}$	$\{Iago, Desdemona\}$
Act 5	$\{Othello, Desdemona\}$	$\{Iago, Desdemona\}$

Table 7.2: Table showing the Unique Personal Extensions of Othello and Iago in the society S for each Act

	Act 1		Act 2		Act 3		Act 4		Act 5	
	$\mu_i(S)$	$\varepsilon_i(S)$	$\mu_i(S)$	$\varepsilon_i(S)$	$\mu_i(S)$	$\varepsilon_i(S)$	$\mu_i(S)$	$\varepsilon_i(S)$	$\mu_i(S)$	$\varepsilon_i(S)$
$i = Othello$	0.0	0.125	0.0	0.25	0.0	0.5	0.0	0.5	0.0	0.187
$i = Iago$	0.0	0.5	1.0	1.0	1.0	1.0	1.0	1.0	0.333	0.395
$i = Cassio$	1.0	0.75	0.0	0.25	0.0	0.25	0.0	0.25	0.333	0.458
$i = Roderigo$	1.0	1.0	1.0	1.0	*na*	*na*	*na*	*na*	0.333	0.833
$i = Desdemona$	1.0	1.0	1.0	1.0	1.0	0.75	1.0	0.75	1.0	1.0
$i = Brabantio$	1.0	1.0	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*

Table 7.3: Table showing the values of μ_i, Expected Trustworthiness, and the $\varepsilon_i(S)$, Coalition Expected Trustworthiness, with regard to the Othello's society S for each Act

	Maximal Trusted Extensions	T-Coalitions
Act 1	$\{Desdemona, Roderigo, Cassio, Brabantio\}$	$\{Desdemona, Brabantion\}$ $\{Othello, Desdemona, Cassio\}$
Act 2	$\{Desdemona, Roderigo, Iago\}$	$\{Roderigo, Iago\}$ $\{Othello, Desdemona\}$ $\{Desdemona, Cassio\}$
Act 3	$\{Desdemona, Iago\}$	$\{Desdemona, Cassio\}$
Act 4	$\{Desdemona, Iago\}$	$\{Desdemona, Cassio\}$
Act 5	$\{Desdemona, Roderigo\}$	$\{Othello, Desdemona\}$ $\{Cassio, Desdemona\}$

Table 7.4: Table showing the Maximal Trusted Extensions and the T-Coalitions in the Othello's society S for each Act after applying the Inference Rule

inference rule. However there is already a direct and mutual trust relationship between Desdemona and Othello. Therefore there is no change in the values of the *Maximal Trusted Relationships*, *Personal Extensions* and *aggregate measures* for the first act.

7.3 ACT 2 Synopsis

Several weeks later in Cyprus, Othello's arrival is expected. But a terrible storm has largely battered and destroyed the Turkish fleet, which no longer poses a threat to Cyprus. Unfortunately there are fears that this same storm drowned Othello as well. Many people praise Othello. Cassio, who has arrived, sings Desdemona's praises. A ship is spotted but it is Desdemona and Iago's, not Othello's. Iago suspects that Cassio loves Desdemona and slyly uses it to his advantage. Iago tells Roderigo that he still has a chance with Desdemona but Cassio whom Desdemona could love is in the way. Killing Cassio (who became Othello's lieutenant instead of Iago) will leave Desdemona to Roderigo, Iago slyly explains.

Othello finally arrives, to everyone's great relief. Iago decides to tell Othello that Cassio is having an affair with Desdemona, so Iago will be rewarded whilst Cassio will be punished. A Herald announces celebration that "our noble general

	Othello's UPE	Iago's UPE
Act 1	$\{Othello, Desdemona, Cassio\}$	$\{Iago, Desdemona, Brabantio\}$
Act 2	$\{Othello, Desdemona\}$	$\{Iago, Desdemona, Roderigo\}$
Act 3	$\{Othello\}$	$\{Iago, Desdemona\}$
Act 4	$\{Othello\}$	$\{Iago, Desdemona\}$
Act 5	$\{Othello, Desdemona\}$	$\{Iago\}$

Table 7.5: Table showing the Unique Personal Extensions of Othello and Iago in the Othello's society S for each Act after applying the Inference Rule

	Act 1		Act 2		Act 3		Act 4		Act 5	
	$\mu_i(S)$	$\varepsilon_i(S)$	$\mu_i(S)$	$\varepsilon_i(S)$	$\mu_i(S)$	$\varepsilon_i(S)$	$\mu_i(S)$	$\varepsilon_i(S)$	$\mu_i(S)$	$\varepsilon_i(S)$
$i = Othello$	0.0	0.125	0.0	0.25	0.0	0.5	0.0	0.5	0.0	0.218
$i = Iago$	0.0	0.5	1.0	1.0	1.0	1.0	1.0	1.0	0.333	0.156
$i = Cassio$	1.0	0.75	0.0	0.125	0.0	0.25	0.0	0.25	0.333	0.437
$i = Roderigo$	1.0	1.0	1.0	1.0	*na*	*na*	*na*	*na*	0.333	0.833
$i = Desdemona$	1.0	1.0	1.0	1.0	1.0	0.75	1.0	0.75	1.0	1.0
$i = Brabantio$	1.0	1.0	*na*	*na*	*na*	*na*	*na*	*na*	*na*	*na*

Table 7.6: Table showing the values of μ_i, Expected Trustworthiness, and the $\varepsilon_i(S)$, Coalition Expected Trustworthiness, with regard to the Othello's society S for each Act after applying the Inference Rule

Othello!" has defeated the Turkish fleet, calling on all to celebrate this great triumph and also to celebrate Othello's "nuptial" or wedding to the fair Desdemona.

Iago learns more of Cassio's high regard for Desdemona and Iago manipulates Cassio into drinking too much since he is certain Cassio will do something he will regret. Iago also tells Roderigo to attack Cassio. This happens, and Cassio wounds Roderigo. Othello is now awake and Cassio's name ruined. Othello, although he loves Cassio, has no choice but to demote him from his position as his lieutenant. Next Iago comforts Cassio by suggesting he speak with Desdemona who could put in a good word for him with Othello. Iago comforts a wounded Roderigo, telling him he has won by ruining Cassio's name. Iago has his wife Emilia ensure Desdemona and Cassio will talk so Othello can see his wife talking with Cassio, allowing Iago to convince Othello that Desdemona is being unfaithful.

7.4 ACT 2 Analysis

In the second act, as shown in Figure 7.3 and 7.4, the distrust and trust relationships between the characters change. Our models reflect these changes.

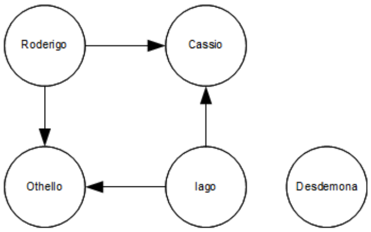

Figure 7.3: ATF for Act 2

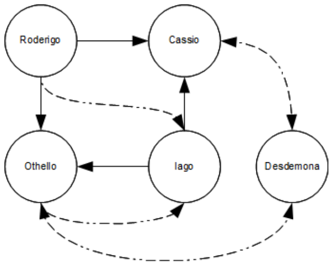

Figure 7.4: ATDF for Act 2

We can observe that now Iago has become distrust free, since Roderigo has regained trust in him. Roderigo now distrusts Cassio, as well as Othello, convinced by Iago that he should attack Cassio to ruin Othello's name.

The *Maximal trusted extension* now includes Iago and excludes Cassio. It is, in fact formed by

$$\{Desdemona, Roderigo, Iago\}.$$

This newly acquired distrust-freeness, boosts Iago's *coalition expected trustworthiness* to 1, while Cassio's decreases to 0.25, since Othello has lost trust in him because of his fight with Roderigo.

Iago's *unique personal extension* now includes Roderigo, having the distrust relationship from Roderigo to Iago being removed.

Othello's *coalition expected trustworthiness* improves slightly, being 0.25 in this act. This is due to the absence of Brabantio in this act. His *unique personal* *extension* does not vary. It still includes Cassio because, even though Othello has lost *trust* in him, due to his fight, Iago has still not convinced Othello of Cassio's betrayal.

The *T-Coalitions* now include the one formed by $\{Roderigo,\ Iago\}$ which reflect the actual cooperation between the two in this second act, even if Iago's motives are not genuine.

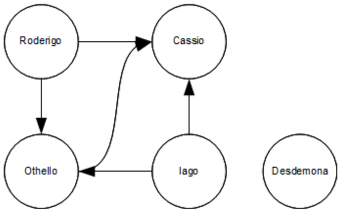

Figure 7.5: ATF for Act 2 after applying the inference rule

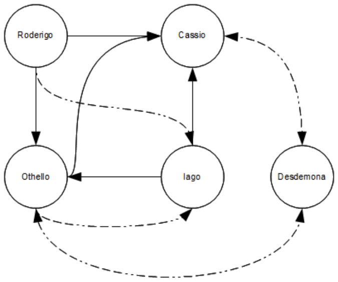

Figure 7.6: ATDF for Act 2 after applying the inference rule

As shown in Figure 7.6, in this act, we can infer a new distrust relationship from Othello versus Cassio, caused by the fact that Othello is trusting Iago and is falling for Iago's lies regarding Cassio. We can also notice that in this act the trust relationship from Othello to Cassio has disappeared due to the incident that has seen Cassio wounding Roderigo. The absence of this trust relation allows now Othello to have doubt regarding Cassio and therefore trusting Iago's words. This is also reflected in Othello's *unique personal extensions* which now contains only Desdemona.

We can also notice a change in the T-Coalition. Othello and Cassio, clearly, are not in the same T-Coalition anymore. Hence, Desdemona, who trusts both, is in two different T-Coalitions, one with Othello and one with Cassio. The events in this act also cause Cassio's *Coalition Expected Trustworthiness* to decrease to 0.125.

7.5 ACT 3 Synopsis

Cassio tells Iago that he has arranged to meet Desdemona, Iago helping Cassio to do this. Iago's wife, Emilia, tells Cassio that Othello would like to reinstate him as his lieutenant but the fact that Cassio's fight is public news prevents Othello from doing this immediately. Emilia tells Cassio that she can arrange a meeting with Desdemona. Some time later, Cassio speaks with a very sympathetic Desdemona who assures him that Othello still very much loves Cassio. Furthermore, Desdemona promises to keep putting in a good word for Cassio until he is again Othello's lieutenant. At a distance, Iago manipulates Othello by first suggesting shock and then hiding his outbursts from Othello. This guarantees Othello's attention, as Iago plants seeds of doubt in Othello's mind about Desdemona's fidelity especially where Cassio is concerned. Iago leaves Othello almost convinced that his wife is having an affair with Cassio.

Othello now complains of a headache to Desdemona, which results in her dropping a strawberry patterned handkerchief, Othello's first gift to her. Emilia picks this up and gives it to Iago who decides the handkerchief could help his manipulation if he ensures Cassio receives it. Iago arranges to place the handkerchief near Cassio's lodgings or home where he is certain to find it and take it as his own, unaware that it is Othello's gift to Desdemona.

A furious Othello returns to Iago, certain his wife is faithful and demanding proof from Iago of Desdemona's infidelity. Reluctantly and hesitantly, Iago tells Othello he saw Cassio wipe his brow with Desdemona's handkerchief. Othello is convinced, cursing his wife and telling Iago who is now promoted to lieutenant to kill Cassio. Othello will deal with Desdemona. Desdemona worries about her missing handkerchief and comments that if she lost it, it could lead Othello doubting her fidelity. Emilia, when asked about Desdemona's lost handkerchief, lies, denying having seen the handkerchief she picked up and gave to Iago. Othello enters; asking Desdemona for the very same handkerchief and Desdemona assures him that the handkerchief is not lost and will be found. Desdemona now tries to change the subject to Cassio, but Othello continually stresses the value the handkerchief has to him, this leading to Othello angrily ordering his wife away.

Cassio arrives, Desdemona telling him that her attempts to help him are not going well. Iago claims total ignorance to the cause of Othello's fury. Cassio gives Othello's handkerchief, which he found, to his suspicious mistress Bianca.

7.6 ACT 4 Synopsis

Iago fans the flames of Othello's distrust and fury with Desdemona's supposed "infidelity" by first suggesting Desdemona shared her bed with Cassio, and then that Desdemona's gesture of giving away the handkerchief is no big deal. But Iago knows exactly how hurtful to Othello this is, giving away this sentimental gift.

Next, Iago suggests to Othello that Cassio will "blab" or gloat to others about his conquest of Desdemona and then tells Othello that Cassio boasted to him that he did indeed sleep with Desdemona. Meeting later with Cassio, Iago cunningly talks to Cassio about Cassio's mistress Bianca, each smile and each gesture made by Cassio infuriating a hidden Othello who thinks Cassio is talking about sleeping with Desdemona. Next Bianca arrives, angrily giving back the handkerchief Cassio gave to her. This infuriates Othello since as Iago puts it, Cassio not only received Othello's handkerchief from his wife but then gave it away to Bianca as if it were worthless. Othello decides to kill Desdemona by strangulation in her bed, Iago's idea. Iago pledges to kill Cassio.

Lodovico (Desdemona's cousin) arrives, announcing that Othello is to return home and Cassio is to be the next Governor of Cyprus. Desdemona's joy for Cassio enrages Othello, leaving Lodovico and Iago to wonder how much Othello seems to have changed and leaving poor Desdemona to wonder how she offended the man she truly loves. Othello questions Emilia as to whether Desdemona was unfaithful to him. Annoyed that Emilia's answers suggest nothing has happened between Desdemona and Cassio, Othello dismisses her comments as those of a simple woman. Othello meets Desdemona, Desdemona becoming increasingly upset with her husband's anger towards her, an anger she cannot understand. Othello eventually reveals to Desdemona that her infidelity is the source of his anger, Desdemona pleading her innocence on deaf ears. Emilia and Desdemona discuss Othello's strange behavior. Emilia is certain some evil fellow has twisted Othello to believe Desdemona has been unfaithful, not realizing that this evil man is her own husband Iago.

We learn that Iago has been pocketing Roderigo's gifts to Desdemona, which never reached her. Fearing Roderigo will learn this, Iago tells Roderigo that Cassio must die since Iago benefits if either man dies. Lodovico tries to calm Othello down. Othello orders Desdemona to bed to await him later, an order Desdemona dutifully obeys out of love for Othello. Emilia notices that Othello is much calmer now and tells Desdemona her bed has been made with her wedding sheets as requested. Desdemona asks to be buried in those same sheets should she die before Emilia, a hint of trouble ahead.

Emilia is forbidden from joining Desdemona in her bedchamber, angering her. Desdemona, depressed, recalls a song (The Willow Song) of a maid who was similarly abused by her husband and sings it. Desdemona talks to Emilia from behind the door about infidelity. Desdemona claims she would not be unfaithful to her husband for all the world while the more cynical and worldly Emilia admits she would, for the right price.

7.7 ACT 3 and 4 Analysis

In the third act, Othello is convinced that Desdemona and Cassio are having an affair, therefore he distrusts them both. Hence, in our model two new *distrust* relationships appear.

Desdemona, although she is now distrusted by Othello, remains in the *maximal trusted extension*, indirectly defended by Iago's distrust toward Othello. Iago still remains distrust-free in the society.

Cassio's *coalition expected trustworthiness* decrease to 0.25, while Iago's reaches 1. This reflects the actual situation in the play. Iago is considered a very loyal friend.

Othello is now alone. He does not trust Cassio or Desdemona anymore. This is shown by the fact that his *unique personal extension* has only himself in it.

Note that now the only *T-Coalition* is formed by $\{Desdemona, Cassio\}$ who,

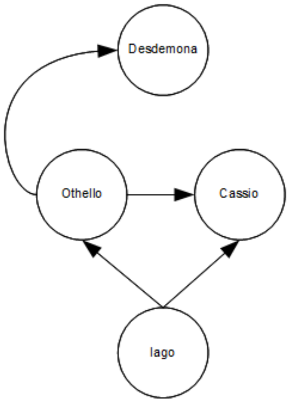

Figure 7.7: ATF for Act 3 and 4

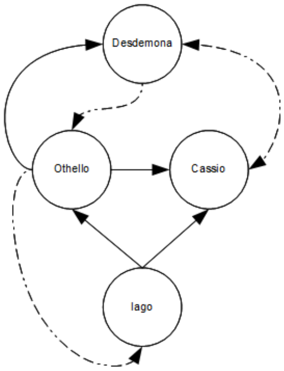

Figure 7.8: ATDF for Act 3 and 4

being still unaware of Iago's plot, keep trusting each other.

Again, in this act, no new *distrust relationship* can be inferred due to the structure of the *trust relationships* among the characters in the society.

For instance, Desdemona trusts Othello even if he doesn't trust her anymore. Othello now despises Cassio. This would cause Desdemona to distrust Cassio as well. However she is trying to help Cassio to regain trust from Othello, therefore there is already a direct and mutual trust relationship between Desdemona and Cassio. Since there is no additional distrust relationship there is no change in the values of the *Maximal Trusted Relationships*, *Personal Extensions* and *aggregate measures* for the this act.

In the fourth act, the *trust* and *distrust* relationships remain unchanged with respect to the third act.

7.8 ACT 5 Synopsis

Iago and Roderigo wait in a street to ambush Cassio. Iago tells Roderigo how to kill him. Iago does not care which ends up dead. Iago is worried about Roderigo's increasing questioning of what happened to jewels that were given to him to pass on to Desdemona. Roderigo attacks Cassio but Cassio wounds Roderigo instead. Iago from behind stabs Cassio, wounding him in the leg. Othello hearing Cassio's cries is pleased, announcing that he too will soon kill (Desdemona). Lodovico and Gratiano (Desdemona's uncle) and Iago reappear, Iago claiming total innocence to Cassio's injuries even though he inflicted them. Seizing Roderigo, Iago stabs and wounds him "in revenge" for wounding his "friend" Cassio. Gratiano and Lodovico tend to Cassio's wound. Bianca, Cassio's mistress arrives, Iago cleverly laying suspicion for Cassio's injuries on his innocent mistress, shifting suspicions from Iago.

Othello enters Desdemona's bedchamber trying to convince himself that he is killing her for her own good. He kisses his still asleep wife one last time. Desdemona awakens, but Othello will still kill her, telling her to pray so her soul will not die when she does. Desdemona again asks what wrong she has committed, Othello telling her that she gave Cassio his handkerchief, by which he means he thinks she had an affair with him. Desdemona pleads her innocence, telling Othello to bring Cassio over to prove she did not give away her handkerchief. Othello says he confessed and is dead, Desdemona's fear and surprise prompting Othello to believe she does care for him.

Othello kills Desdemona.

Emilia, banging on the door outside, cannot stop this. Later Emilia is let in, revealing Iago has killed Roderigo and Desdemona who was thought dead, murmurs her last breaths but loyally does not say Othello killed her. Othello tells Emilia he killed her and Emilia despite Iago's attempts to remove her reveals the truth about the handkerchief; she found it, and then gave it to Iago. Iago, now in trouble, stabs his wife Emilia and escapes. Emilia dies, singing the "Willow Song" before criticizing Othello for killing his loving wife.

Lodovico, Cassio and the now captured prisoner Iago soon appear, Othello stabbing Iago but not killing him before having his sword removed. Lodovico is disappointed that Othello, a man so honorable has reverted to acting like a slave. Othello tries to argue that killing his wife was a noble action but it falls on deaf ears. Lodovico learns that Othello and Iago plotted Cassio's death. Lodovico reveals letters in the dead Roderigo's pocket proving Cassio was to be killed by Roderigo. Iago proudly confirms that Cassio did find the handkerchief in his bedchamber because Iago placed it there to be found. Othello, realizing what he has done, kills himself with a concealed weapon and lies himself on top of his wife. Cassio is placed in charge of Iago and Lodovico leaves to discuss this sad matter with others abroad.

7.9 ACT 5 Analysis

In the end of the final act, the truth is finally revealed. Roderigo finally understands that Iago has used him. Othello realises that he has been a fool and victim of Iago's plot. This is shown by the new *distrust* relationships from Roderigo and Othello to Iago.

Desdemona is now dead and Cassio realises that Othello was planning to kill him too. Cassio now distrusts Othello and Iago. On the other side, Othello now trusts Cassio again, realizing that Cassio has always been loyal to him.

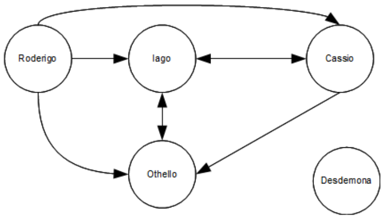

Figure 7.9: ATF for Act 5

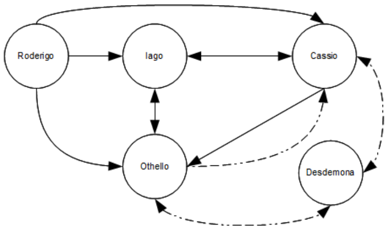

Figure 7.10: ATDF for Act 5

The last act changes Iago's situation drastically. His *coalition expected trustworthiness* drops from 1 to 0.3. His *unique personal extension* includes only Desdemona who died before realising she was part of the Iago's plot.

Othello's situation remains bad. His *coalition expected trustworthiness* is very low, 0.18. His *unique personal extension* includes Desdemona, who, even before dying, refused to betray her beloved husband by revealing that he was her killer.

Othello's renewed trust in Desdemona means that there is now another *T-Coalition* formed by the two, since Desdemona never stopped trusting Othello.

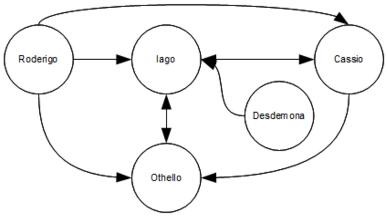

Figure 7.11: ATF for Act 5 after applying the inference rule

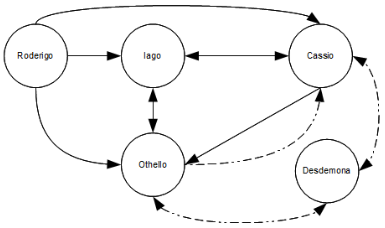

Figure 7.12: ATDF for Act 5 after applying the inference rule

In this act, we can infer a new *distrust* relationship from Desdemona to Iago. This means that Iago is now distrusted by everyone in the society. This is the first time since the first act, where he was distrusted by Roderigo, that he does no longer appear in the *Maximal Trusted Extension*. His *coalition expected trustworthiness* drops even further from 0.3 to 0.15. His *unique personal extension* does not include any other character in the society. Iago is now completely alone and distrusted.

7.10 Summary

In this chapter we have used the famous tragedy by Shakespeare, Othello, to give an actual representation and analysis of the models proposed.

The implementation of the ATF framework and the functionality to compute the *Maximal Trusted Extensions*, *Personal Extensions* and all the *aggregate measures* helped extract the data for the illustrative tables.

The analysis of the Othello scenario has shown that the models can successfully reproduce combinations of trust and distrust relationships for realistic situations. The changes in the relationships are reflected in our models and so are the values of the aggregate measures, working as a mirror of the society.

Chapter 8

Conclusions and Future Work

8.1 Conclusions

8.1.1 Building and Using Social Structures

In the first part of this thesis we have described a method to allow agents to use information about the society they live in to make decisions about trust.

We have seen that in trust and reputation research using multiple sources of information is crucial to allow a robust trust model to function effectively. The combination of various methods and sources of information, including those related to social relations, allows the agent to calculate trust and reputation values at different stages of knowledge of the society. Some trust and reputation systems [63] have recognized the value of using information about social relationships to support trust decisions. However, we have explained that little work has been done to properly *evaluate* and *validate* this added value. Moreover, the issue of how a social structure, representing these social relations between agents in the society, is *created* at the start, and how it *evolves*, does not appear to have been considered in the literature.

Therefore, we have presented a technique for agents to build a social network representation of their local environment. As the agents interact with other agents they gather information about interactions and relationships in order to build the network of agents and to better understand their social environment. We have shown that we are able to identify 3 types of social relationships:

- Competition (COMP). This is the type of relation found between two agents that pursue the same goals and need the same (usually scarce) resources. This kind of relation could appear between two sellers that sell the same product or between two buyers that need the same product.
- Cooperation (COOP). This relation implies the exchange of sincere information between the agents and some kind of predisposition to help each other if possible. In other words, we assume that two agents cannot have at the same time a competitive and a cooperative relation.
- Trade (TRD). This type of relation is compatible either with cooperative or competitive relation. It reflects the existence of some commercial transactions between two agents, without a distinctive competitive or collaborative behaviour.

According to the type of relation linking two agents, they can make different decisions when it comes to trust-related situations.

We have also shown empirical evidence that a trust model enhanced with a social structure representation, used to gather additional information to select trustworthy agents for an agent's interactions, can improve the trust model's performance.

8.1.2 A Framework for Trust and Distrust

In the second part of this thesis, we have introduced an abstract framework that takes trust and distrust into account for coalition formation. Coalition formation, the process by which a group of software agents come together and agree to coordinate and cooperate in the performance of a set of tasks, is an important form of interaction in multi-agent systems. Such coalitions can improve the performance of the individual agents or the whole system, especially when tasks cannot be performed by a single agent, or when a group of agents would perform the tasks more efficiently.

With a relatively small number of exceptions mentioned in Chapter 2, existing models of coalition formation do not consider trust [17, 41] and in more general models [56, 38], individual agents use information about reputation and trust to rank agents according to their level of trustworthiness. We have explained that these models lack a *global* view because they only consider the trust binding the agent starting the coalition and the agents receiving the request to join the coalition.

We have addressed these limitations through the definition of an abstract framework that allows agents to form distrust-free coalitions. We have formulated several notions of mutually trusting coalitions. We have also presented techniques for how the information presented in our distrust model can be aggregated to produce individual measures to evaluate the trustworthiness of the agent with respect to the whole society or to a particular coalition. We also have presented a way to combine the trust and distrust relationships to form coalitions which are still distrust-free. Moreover, the analysis of the Othello's scenario shown that the models can successfully reproduce combinations of trust and distrust relationships for realistic situations. The changes in the relationships are reflected in our models and so are the values of the aggregate measures, working as a mirror of the society.

8.2 Future Work

We have divided the future work in two parts. The future work related with building and using social structures in agent societies and the future work related with the abstract framework for trust and distrust.

8.2.1 Building and Using Social Structures

Although our approach makes a number of advances to the state of the art, there are still a number of ways in which this work can be further extended. Here we propose a short list of topics we think would improve our approach:

- Our quantitative results and conclusions are based on observations. As introduced in Section 4.4, for certain measures the results for the social agent simply dominate those for the non-social agent under each scenario. For the other measures, more subtle conclusions were drawn. Therefore, our results would benefit from a more systematic analysis using an in-depth statistical toolbox on the results for the various scenarios, and investigating which parameters are interesting to analyse further.

- Having considered the remarks made on the Agent ART testbed, similar testbed could be considered as an alternative, such as the "Trading Agent Competition" [3]. However these frameworks present similar issues as the Agent ART testbed. In the Trading Agent Competition, agents will have to demonstrate ability in dealing with decisions about variations in customer demand and availability of supplies. Again, we feel that these factors could contribute negatively to the agent performance with regard with trust. Therefore, we feel that it would be more useful to set up a simpler testbed, where the factors involved in the success of an interaction are only the ones related with the trust decisions. Similar test sets have been used for the well-known Prisoner's Dilemma Competition by [8] and by the more recent [15]. Furthermore, additional testing of our approach would validate further the general nature of the method.

- In the computation of the distance measure, one overall value for all the different eras is calculated. Therefore it is difficult to separate particular eras where expertises of the two agents diverge more than those where the difference is minor. This could lead to the computation of a value that does not represent accurately the level of conflict between the agents. Recent work [63, 40] in trust and reputation models consider trust as a multi-facet concept, hence in our scenario, different eras can be viewed as different perspectives.

- Moreover, in future works, it might be interesting to consider that, for a same amount of difference in expertise, the behaviours (cooperative/competitive) of the agents might differ according to their current goal, whether it is gathering information, or providing it or selecting a partner for a certain interaction.

- The way the social structure is updated every time new information is acquired could be refined by using link prediction techniques [43] and path analysis [50] and explore other measures of similarity.

- Finally, in our work we do not consider the possibility that the agents may lie in the expertise value they provide to others. Therefore, it would be useful to investigate measures to normalize the values received to account for the possibility of agents lying.

8.2.2 A Framework for Trust and Distrust

Coalition stability is a crucial issue. Stability is the motivation of an agent's refusal to break from the original coalition and form a new one. We have presented several notions of mutually trusting coalitions that we believe improve current coalition formation techniques based on trust. However, there are aspects that need improvement.

- We have at various times been talking about the notion of *stability* in *coalition formation*. As mentioned in 4.4 there is the cooperative game theory literature on stability in coalition formation – see [54] for details. The best-known notion of stability in cooperative game theory is the *core*. This solution concept considers a coalition to be stable if no subset of the coalition has any rational incentive to defect from the coalition, in the sense that they could earn more for themselves by defecting. In our view, a coalition is stable if no agent has any rational incentive to distrust any of the members. Future work might consider examining the role of our distrust models in coalition formation in more detail, perhaps in the context of coalition formation algorithms such as those recently proposed within the MAS community (see, e.g., [64]).

- There are two factors determining an agent's trust in a coalition, one is that the agent must possess necessary competence for a particular task, the other is that all the agents involved in a coalition must behave honestly and diligently [56]. In this work, we have abstracted from the competence skills required to accomplish a particular task, but we concentrated on the belief that the agent needs to have that its fellow coalition members will perform the task assigned without defecting. An intersting development would be the ability to combine information about skills and competence together with information about trust during the computation of the coalition. This will allow for different perspectives to be analysed. By introducing this measure, we aim to make the framework more general, in order for it to consider features other than trust.

- Our work is based on a boolean notion of trust. However in real scenarios, an agent has neither a complete trust or no trust at all in another agent. There could be various degree of trust in between. For example, in the Othello scenario, in Act 2, Othello's trust in Cassio changes during the act. We have represented this simply by removing the trust relationship between the two. However, it's clear for the story, that Othello feels sorry of having to demote Cassio. Hence, he still has some degree of trust toward Cassio. Therefore being able to add degrees of trust or distrust to our model would allow to frame more specific situations and cater for more detailed information. Perhaps, the work carried out with regard to weights in argumentation systems [27] could also be applied in our context.

- Moreover, we assume that the agents are willing to share their information about the trust they have in other agents. It would be interesting to devise some form of incentives for the agents to do so.

- And finally, agents are assumed to be constantly updating their information as they learn more about the others, which argues for adding a time element to the framework to make it dynamic. The concept of transitivity and therefore our trust-induced inference rules would be affected by this dynamism, as such the model might need to keep track of induced relationships so that they could be backtracked when changes in the society occur. This would allow the agent to form coalitions when it is advantageous to them or to better select trustworthy and competent agents to work together, allowing for a greater stability in the society.

This point also relates with having the possibility to define degrees of trust and distrust. Being able to monitor changes in these values could give agents important insight into the dynamics of the social relationships in the society and help them to revise their own beliefs when these are based, for example, on inferred relationships.

Bibliography

[1] Othello summary. `http://absoluteshakespeare.com/guides/summaries/othello/othello_summary.htm`, 2005.

[2] *FilmTrust: movie recommendations using trust in web-based social networks*, volume 1, 2006.

[3] J. Collins A, R. Arunachalam, N. Sadeh B, J. Eriksson, N. Finne, and S. Janson C. The supply chain management game for the 2007 trading agent competition, 2006, Carnegie-Mellon University.

[4] A. Abdul-Rahman and S. Hailes. Supporting trust in virtual communities. In *Proceedings of the 33rd Hawaii International Conference on System Sciences*, HICSS '00, pages 6007–, 2000.

[5] L. Amgoud, C. Cayrol, and M. C. Lagasquie-Schiex. On the bipolarity in argumentation frameworks, 2004, IRIT-UPS.

[6] R. Ashri, S. D. Ramchurn, J. Sabater, M. Luck, and N. R. Jennings. Trust evaluation through relationship analysis. In *AAMAS '05: Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, pages 1005–1011, New York, NY, USA, 2005. ACM.

[7] K. Atchariyachanvanich and N. Sonehara. Trust perception in internet shopping: comparative study of customers in japan and south korea. In *ICEC '08: Proceedings of the 10th international conference on Electronic commerce*, pages 1–8, New York, NY, USA, 2008. ACM.

[8] R. Axelrod. *The Evolution of Cooperation*. Basic Books, 1984.

[9] M. Bacharach and D. Gambetta. Trust as type detection. *Trust and deception in virtual societies*, pages 1–26, 2001.

[10] B.. Barber. *Science and the social order*. Allen and Unwin, London :, 1953.

[11] K. S. Barber and J. Kim. Belief revision process based on trust: Agents evaluating reputation of information sources. In *Trust in Cyber-societies, LNAI 2246*, pages 73–82. Springer, 2001.

[12] K. S. Barber and J. Kim. Soft security: isolating unreliable agents from society. In *Proceedings of the 2002 international conference on Trust, reputation, and security: theories and practice*, AAMAS'02, pages 224–233, Berlin, Heidelberg, 2003. Springer-Verlag.

[13] P. Baroni and M. Giacomin. Solving semantic problems with odd-length cycles in argumentation. In *ECSQARU*, pages 440–451, 2003.

[14] P Bateson. *The biological evolution of cooperation and trust*, pages 14–30. Basil Blackwell, 1988.

[15] B. Beaufils, J. Delahaye, and P. Mathieu. Complete classes of strategies for the classical iterated prisoner's dilemma. In *Evolutionary Programming*, pages 33–41, 1998.

[16] A. Biswas, S. Sen, S. Debnath, I. Sen, and I. Debnath. Limiting deception in groups of social agents. *Applied Artificial Intelligence*, 14:200–0, 2000.

[17] S. Breban and J. Vassileva. Long-term coalitions for the electronic marketplace. In *Proceedings of the E-Commerce Applications Workshop, Canadian AI Conference*, 2001.

[18] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. In *Seventh International World-Wide Web Conference (WWW 1998)*, pages 107–117, 1998.

[19] B.Yu and M. P. Singh. Distributed reputation management for electronic commerce, 2002, Department of Computer Science, North Carolina State University.

[20] C. Castelfranchi and R. Falcone. Principles of trust for mas: cognitive anatomy, social importance, and quantification. In *Principles of trust for*

MAS: cognitive anatomy, social importance, and quantification, pages 72–79, 1998.

[21] C. Castelfranchi and R. Falcone. *Trust Theory: A Socio-Cognitive and Computational Model*. Wiley Series in Agent Technology. John Wiley and Sons Ltd, Chichester, 2010.

[22] C. Castelfranchi and Y. Tan. The role of trust and deception in virtual societies. In *in Proceedings of the 34 th Annual Hawaii International Conference on Systems Science, Maui, Hawaii, IEEE Computer*, pages 7011–7018. Society Press, 2001.

[23] M. Chapman, G. Tyson, K. Atkinson, M. Luck, and P. McBurney. Social networking and information diffusion in automated markets. In *Proceedings of the Joint Workshop on Trading Agent Design and Analysis (TADA) and Agent-Mediated Electronic Commerce (AMEC)*, pages 89–102, 2012.

[24] B. Christianson and W. Harbison. Why isn't trust transitive? In Mark Lomas, editor, *Security Protocols*, volume 1189 of *Lecture Notes in Computer Science*, pages 171–176. 1997.

[25] M. Deutsch. *Cooperation and trust: some theoretical notes*. Nebraska University Press, 1962.

[26] P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and *n*-person games. *AI*, 77:321–357, 1995.

[27] P. Dunne, A. Hunter, P. McBurney, S. Parsons, and M. Wooldridge. Weighted argument systems: Basic definitions, algorithms, and complexity results. *Artificial Intelligence*, 175(2):457–486, 2011.

[28] P. E. Dunne and T J. M. Bench-Capon. Coherence in finite argument systems. *Artif. Intell.*, 141(1-2):187–203, October 2002.

[29] P. E. Dunne and M. Wooldridge. Complexity of abstract argumentation. In Guillermo Simari and Iyad Rahwan, editors, *Argumentation in Artificial Intelligence*, pages 85–104. Springer US, 2009.

[30] B. Esfandiari, S. Chandrasekharan, and Sanjay Ch. On how agents make friends: Mechanisms for trust acquisition. In *Proceedings of the Fourth Workshop on Deception, Fraud and Trust in Agent Societies*, pages 27–34, 2001.

[31] R. Falcone and C. Castelfranchi. Transitivity in trust. a discussed property. WOA, 2010.

[32] D. S. Felsenthal and M. Machover. *The Measurement of Voting Power*. Edward Elgar: Cheltenham, UK, 1998.

[33] K. K. Fullam, T. B. Klos, G. Muller, J. Sabater, A. Schlosser, Z. Topol, K. S. Barber, J. S. Rosenschein, L. Vercouter, and M. Voss. A specification of the agent reputation and trust (art) testbed: experimentation and competition for trust in agent societies. In *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, AAMAS '05, pages 512–518, New York, NY, USA, 2005. ACM.

[34] K. K. Fullam, Tomas B. Klos, G. Muller, J. Sabater-Mir, Z. Topol, K. S. Barber, J. Rosenschein, and L. Vercouter. The agent reputation and trust (art) testbed architecture. In *Proceeding of the 2005 conference on Artificial Intelligence Research and Development*, pages 389–396, Amsterdam, The Netherlands, The Netherlands, 2005. IOS Press.

[35] J.. Galaskiewicz and S. Wasserman. *Advances in social network analysis : research in the social and behavioral sciences / Stanley Wasserman, Joseph Galaskiewicz, editors*. Sage Publications, Thousand Oaks, Calif. :, 1994.

[36] L Gasser. Social conceptions of knowledge and action: Dai foundations and open systems semantics. *Artificial Intelligence*, pages 107–138, 1991.

[37] J. Golbeck and J. Hendler. Inferring binary trust relationships in web-based social networks. *ACM Trans. Internet Technol.*, pages 497–529, 2006.

[38] N. Griffiths and M. Luck. Coalition formation through motivation and trust. In *Proceedings of the second international joint conference on Autonomous agents and multiagent systems*, AAMAS '03, pages 17–24, New York, NY, USA, 2003. ACM.

[39] B. Horling and V. Lesser. A survey of multi-agent organizational paradigms. *The Knowledge Engineering Review*, pages 281–316, 2004.

[40] T. D. Huynh, N. Jennings, and N. R. Shadbolt. An integrated trust and reputation model for open multi-agent systems. *Autonomous Agents and Multi-Agent Systems*, pages 119–154, 2006.

[41] G. Lei, W. Xiaolin, and Z. Guangzhou. Trust-based optimal workplace coalition generation. In *Information Engineering and Computer Science, 2009. ICIECS 2009. International Conference on*, pages 1 – 4, 2009.

[42] X. Li, Z. Han, and C. Shen. Transitive trust to executables generated during runtime. In *Proceedings of the Second International Conference on Innovative Computing, Informatio and Control*, ICICIC '07, pages 518–, Washington, DC, USA, 2007. IEEE Computer Society.

[43] D. Liben-Nowell and J. Kleinberg. The link prediction problem for social networks. In *CIKM '03: Proceedings of the twelfth international conference on Information and knowledge management*, pages 556–559, New York, NY, USA, 2003. ACM.

[44] N. Luhmann. *Trust and Power*. Wiley, 1979.

[45] S. P. Marsh. Formalising trust as a computational concept. Technical report, Department of Computing Science and Mathematics, University of Stirling, 1994.

[46] Y. Matsui and T. Matsui. Np-completeness for calculating power indices of weighted majority games. *Theoretical Computer Science*, pages 98–01, 1998.

[47] R. Mukherjee, B. Banerjee, and S. Sen. Learning mutual trust. In Rino Falcone, Munindar Singh, and Yao-Hua Tan, editors, *Trust in Cyber-societies*, volume 2246 of *Lecture Notes in Computer Science*, pages 145–158. Springer Berlin / Heidelberg, 2001.

[48] G. Muller and L. Vercouter. Decentralized monitoring of agent communications with a reputation model. In *Trusting Agents for Trusting Electronic Societies*, pages 144–161, 2004.

[49] J. Murillo and V. Munoz. Agent uno: winner in the 2nd spanish art competition. *Inteligencia artificial: Revista Iberoamericana de Inteligencia Artificial*, pages 19–27, 2008.

[50] S. Wasserman P. J. Carrington, J. Scott. *Models and Methods in Social Network Analysis (Structural Analysis in the Social Sciences).* Cambridge University Press, February 2005.

[51] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. Technical Report 1999-66, Stanford InfoLab, November 1999.

[52] S. Parsons, Y. Tang, E. Sklar, P. McBurney, and K. Cai. Argumentation-based reasoning in agents with varying degrees of trust. In *The 10th International Conference on Autonomous Agents and Multiagent Systems - Volume 2*, AAMAS '11, pages 879–886, Richland, SC, 2011. International Foundation for Autonomous Agents and Multiagent Systems.

[53] J. Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1988.

[54] B. Peleg and P. Sudholter. *Introduction to the Theory of Cooperative Games (second edition).* SV, 2002.

[55] J. L. Pollock. Defeasible reasoning with variable degrees of justification. *Artificial Intelligence*, 133(12):233 – 282, 2001.

[56] Zhou Q., Wang C., and Xie J. Core: A trust model for agent coalition formation. In *Natural Computation, 2009. ICNC '09. Fifth International Conference on*, volume 5, pages 541 –545, 2009.

[57] W. Quattrociocchi, M. Paolucci, and R. Conte. Reputation and uncertainty reduction: Simulating partner selection. In Rino Falcone, Suzanne Barber, Jordi Sabater-Mir, and Munindar Singh, editors, *Trust in Agent Societies*. Springer Berlin / Heidelberg, 2008.

[58] I. Rahwan and G. R. Simari, editors. *Argumentation in Artificial Intelligence*. SV, 2009.

[59] S. D. Ramchurn, D. Huynh, and N. R. Jennings. Trust in multi-agent systems. *Knowl. Eng. Rev.*, pages 1–25, 2004.

[60] P. Resnick and R. Zeckhauser. Trust among strangers in internet transactions: Empirical analysis of ebay's reputation system. In *The Economics of the Internet and E-Commerce*, volume 11 of *Advances in Applied Microeconomics*, pages 127–15. Elsevier Science, 2002.

[61] Stuart J. Russell and Peter Norvig. *Artificial intelligence - a modern approach: the intelligent agent book*. Prentice Hall series in artificial intelligence. Prentice Hall, 1995.

[62] E. Sklar S. Parsons and P. McBurney. Using argumentation to reason with and about trust. In *Eighth International Workshop on Argumentation in Multi-Agent Systems (ArgMAS 2011), Taipei, Taiwan*, pages 89–102, May 2011.

[63] J. Sabater and C. Sierra. Reputation and social network analysis in multi-agent systems. In *AAMAS '02: Proceedings of the first international joint conference on Autonomous agents and multiagent systems*, pages 475–482, New York, NY, USA, 2002. ACM.

[64] T. Sandholm, K. Larson, M. Andersson, O. Shehory, and F. Tohmé. Coalition structure generation with worst case guarantees. *AI*, 111(1–2):pages 209–238, 1999.

[65] M. Schillo, P. Funk, I. Stadtwald, and M. Rovatsos. Using trust for detecting deceitful agents in artificial societies. *Applied Artificial Intelligence*, pages 825–848, 2000.

[66] K. Takcs, B. Janky, and A. Flache. Collective action and network change. *Social Networks*, volume 30:pages 177–189, 2008.

[67] W. T. Teacy, J. Patel, N. R. Jennings, and M. Luck. Travos: Trust and reputation in the context of inaccurate information sources. *Autonomous Agents and Multi-Agent Systems*, 12(2):pages 183–198, 2006.

[68] B. Verheij. Two approaches to dialectical argumentation: Admissible sets and argumentation stages. In *Proceedings of the Eighth Dutch Conference on Artificial Intelligence*, NAIC'96, pages 357–368, 1996.

[69] W.D. Wallis. *A Beginner's Guide To Graph Theory*. Springer, June 2007.

[70] Michael Wooldridge. *An Introduction to Multiagent Systems*. Wiley, 2. edition, 2009.

[71] D. Wu and Y. Sun. The emergence of trust in multi-agent bidding: A computational approach. In *Proceedings of the 34th Annual Hawaii International Conference on System Sciences (HICSS-34)-Volume 1*, HICSS '01, pages 1041–. IEEE Computer Society, 2001.

[72] Y. Wu and M.W.A. Caminada. A labelling-based justification status of arguments. *Studies in Logic*, 3(4):pages 12–29, 2010.

[73] R. R. Yager, J. Kacprzyk, and M. Fedrizzi, editors. *Advances in the Dempster-Shafer theory of evidence*. John Wiley & Sons, Inc., 1994.

[74] G. Zacharia, A. Moukas, and P. Maes. Collaborative reputation mechanisms in electronic marketplaces. In *HICSS '99: Proceedings of the Thirty-second Annual Hawaii International Conference on System Sciences-Volume 8*, page 8026, Washington, DC, USA, 1999. IEEE Computer Society.