Organizational Survival and the Emergence of Collaboration Networks: a Multi-Agent Approach

Pedro José Ramos Moreira de Campos

Tese de Doutoramento em Ciências Empresariais

Orientada por:

Pavel B. Brazdil

Maria Paula Brito

Faculdade de Economia

Universidade do Porto

2007

To Isabel, Té, and Miguel

To my parents

Biographic Note

Pedro José Ramos Moreira de Campos was born in Porto, Portugal, on June 11, 1965. He finished his high-school education and started undergraduate studies at Portucalense University, Porto, in 1985. In 1990 he received a degree on Mathematics applied to Statistics and Operations Research. He initiated graduate studies in 1994, at FEP (Faculty of Economics at University of Porto) and concluded *Provas de Aptidão Pedagógica e Capacidade Científica,* having presented the dissertation "*Elements of Mathematical Demography - with applications to the recent situation of Portuguese Population",* and a pedagogical work on "*Sampling Theory*". He has been a Ph.D student in Management (FEP) since 2003.

He has also been a lecturer at FEP since 1991 and a collaborator of the Portuguese National Statistical Institute (INE) since 1990, where he has been responsible for the Local Dissemination Department (2001-2004) and participated in population studies. He is also a consultant in the area of statistics and market studies.

Acknowledgements

I thank my supervisors, Professors Pavel Brazdil and Paula Brito for their support throughout this work. Their encouragement, dedication and guidance were determinant for my research. Sincere gratitude to Professor Dalila Fontes, from the Management Group in the Faculty of Economics of Porto (FEP), for her support in monitoring the thesis.

I want to express my gratefulness to my wife, Isabel, whose contributes in the microeconomic foundations were of great importance; to my colleague Kaustubh Patil who kindly helped me in the programming of some libraries in the R package; to Professors Mário Rui Silva, Sandra Silva, Catarina Roseira, L. Mota de Castro and Carlos Brito (FEP) and Luís Araújo (University of Lancaster), for their helpful suggestions.

I would also like to acknowledge the helpful support provided by LIADD and FEP, particularly for supporting the presentation of preliminary essays international conferences. Sincere gratitude to my colleagues in LIAAD (Alípio, João, Carlos, Rui, Rodolfo, Rita, Pedro), and FEP (Adelaide, Fernanda, Jorge, and many others), whose companionship is most truthfully acknowledged.

A final acknowledge for Professor Paulo Gomes, whose enthusiasm and passion have been an inspiration for my research and academic experiences all over the years.

Abstract

The present work deals with interactions between firms. Literature that inspired our work comes from organizational ecologists, evolutionary economics and Agent-Based Computational Economics. Our general belief is that organizations are adaptive at different levels of analysis: individually, or as a group. We dedicate a special attention to the network formation and regard networks as new forms of organizations.

In order to understand these interactions and the different mechanisms of adaptation, that include learning and evolution, two studies have been developed in this work: *CASOS* and *NetOrg*. Our main objective with *CASOS – Cellular Automata System for Organizational Survival* - is to analyse the effects of a set of parameters (organizational density, size and age) in the founding and in the mortality of organizations. The best combination of these parameters will produce the best solution for the simulation. The cellular automata approach is embedded with a Genetic Algorithm that calibrates the parameters in order to validate the final solution with real data coming from the Portuguese Industry. We have observed similar tendencies between simulated and real data that represent the evolution of the number of industries in this region. We used survival analysis techniques that also confirm the capability of the model to reproduce the reality: the effect of the size on the mortality of the firms is negative and statistically significant.

NetOrg – Adaptive Networks of Organizations – is a Multi-Agent framework that aims at analyzing the dynamics of organizational survival in cooperation networks. Firms cooperate horizontally (in the same market) or vertically with other firms. Cooperation decisions involve cognitive and microeconomic modelling. To validate our approach, we considered some evidences from three industrial real life examples: Automobile manufacturing, Textile and e-Marketplaces. We observed that organizations increase their stock of knowledge due to the spread of innovation, and that this is a direct consequence of the fact that organizations are linked to networks. The network is a mechanism undergoing adaptation, where the internal synergies shape the course of evolution.

Table of Contents

List of Tables

List of Figures

“Einstein has said (...) the supreme task is to arrive at those universal laws from which the cosmos can be built up by pure deduction. There is no logical path to these laws”.

Robert M. Pirsig,
in “Zen and the Art of Motorcycle Maintenance
- An inquiry into values”

1. Introduction

Edgar Morin (Morin, 1977) once wrote that an "*organization can be defined as an arrangement of relationships between components or individuals.*" These elements or individuals become the components of a whole (a unit or a system) endowed with qualities not apprehended at the level of the components or individuals.

This description can be applied to any entities defined as components or individuals. In organizational studies, this definition may refer to **components of firms**[1], or **firms**, or even **groups of firms**. In other words, Morin's definition of *organization* considers and includes several stages of analysis that can be interrelated.

The present work deals with the interconnection between these different levels of analysis. Our general belief is that organizations are adaptive at different levels of analysis: individually, or as a group. We dedicate a special attention to the network formation and regard networks as new forms of organizations.

[1] In this work we use indistinctively the terms "firm" and "organization".

Literature that inspired our work comes from organizational ecologists and evolutionary economics. Different levels of analysis were taken into account as organizational scientists used different perspectives to study firms.

During the last three decades, organizational ecologists have developed theories with the aim of understanding the forces that shape the structures of organizations over long time spans. Their scope of analysis is the study of organizational populations as a whole, i.e., ecological concern was most heavily concentrated on the probability of survival of the organizational population, rather than the individual organization.[2]

The organizational competences (Penrose, 1959), the relationships involving organizations, and microeconomic modelling were introduced in organization studies. The field of evolutionary economics (Nelson and Winter, 1982) emerged and, therefore, the possibility of extending individual attributes had a positive effect on the study of inter-organizational networks, a form of interaction among firms through the definition of inter-organizational relations.

Within this perspective, we have introduced inter-organizational *collaboration* or *cooperation*[3]. Collaborative strategies encourage firms to specialize in more critical activities where they can be more proficient letting other activities to be accomplished by other members of the network. Collaboration can be understood as the complementary action between the elements of a network, namely in terms of the efforts towards technological innovation.

The analysis of inter-organizational networks stands simultaneously at the global (population) and local (individual) levels. Both levels of analysis are important and complementary. On one hand, the study of a population at its global level is incomplete.

---

[2] Special attention will be given to the work of Stinchcombe (1965), Hannan and Freeman (1977, 1984), Freeman (1982, 1997), Carroll and Hannan (1989, 1992), among other authors. A. Stinchcombe studied the impact of social conditions as way to introduce the environment on the development of organizational structure. Subsequently, M. Hannan, J. Freeman and G. Carroll, among others, analysed the effects of the environment on organizational structure with a deeper focus on ecological issues as a metaphor to study organizations.

[3] *Collaboration* and *cooperation* will be used indifferently, although these terms have different meanings, as we will see later on.

On the other hand, the analysis of interactions, only at the individual level, may also lead to phenomenon misunderstanding. Usually, mainstream economics assume that the market is in equilibrium and that the individual preferences are constant. Thus, the analysis of mixed levels may constitute the answer to some of these problems.

The links between *individual* and *aggregate* levels are essential to understand global phenomena. According to Axelrod (1997), the large-effects of complex locally interacting individuals lead to the appearance of *emergent* properties at the aggregate global level, i.e., the level of the population.

Organizations interact with the environment and with other organizations, and these interactions constitute an important way of learning and evolution. To overcome the problems that they face during their existence, organizations must certainly adopt survival strategies, both individually and in group.
The way in which organizations adopt survival strategies by network formation to avoid failure constitutes the main motivation of our work. Therefore, the main question to be answered is:

- Are networks of organizations adaptive in a comparable way as organizations are adaptive?

The question of adaptation is an issue that can be studied simultaneously at individual and at population level. At individual level, there is evidence that organizations adapt themselves to environmental pressures by replacing less effective competences with more effective ones. Similarly, the organizational population changes as new members holding more favourable competences drive into failure members holding less effective competences.

It is possible, though, to build an overall view of an adaptation process that is simultaneously individual and collective. Ferber (1999) affirms that adaptation is either a consequence of learning, as an individual characteristic of the agent, or a consequence of evolution. In this last situation, adaptation is seen as a collective process bringing reproductive mechanisms into play.

The specific questions we want to answer are connected with density, survival and collaboration among firms and can be summarized as follows:

- What are the effects of organizational density in the survival of organizations?
- What are the effects of organizational density in the survival of networks?
- What are *the effects of organizational density* on the survival of networks?
- How does cooperation motivate the formation of networks?
- What is the effect of individual learning in the formation of networks?
- Does the network shape matter? In other words, has the shape of the network some kind of influence in its performance?
- Does duration of relationships matters?
- What is the impact of the distance on the type of relationship (vertical/ horizontal) in a network?
- How does knowledge transmit through networks?
- How do cognitive capabilities, that are available in the organizations, intervene in the formation and success of the networks?

In order to understand the interaction between different mechanisms of adaptation, that includes learning and evolution, we have created a simulated environment. Simulation permits to simplify the world, and is a well-recognized way of understanding it. It provides tools for implementing different scenarios, by constructing a virtual economic world with the appropriate parameterisation.

Emergent behaviour of aggregate variables is then captured and the parameters of the model can be reformulated in order to simulate different scenarios related to different socio-economic perspectives.

The simulation of individual agents and the consequent analysis of aggregate behaviour produced from their interaction may be achieved using, for example, Agent-Based modelling, or Cellular Automata. The work benefits from the contributions of Agent-Based Computational Economics (ACE), a recent field of analysis, that aims at "growing economies from the bottom up" (Tesfatsion, 2006). Economies are complicated systems encompassing micro behaviours, interaction patterns, and global regularities.

Two studies have been developed in this work: *CASOS* and *NetOrg*. Our main objective with *CASOS – Cellular Automata System for Organizational Survival* - is to analyse the effects of certain parameters in the founding and in the mortality of organizations. The best combination of these parameters will produce the best solution for the internal parameters of the simulation. The cellular automata approach is embedded with a Genetic Algorithm that calibrates the parameters in order to validate the final solution with real data coming from the Portuguese Industry.

NetOrg – Adaptive Networks of Organizations – is a Multi-Agent framework that aims at analyzing the dynamics of organizational survival in cooperation networks. Firms cooperate horizontally (in the same market) or vertically with other firms that belong to the supply chain. Cooperation decisions involve cognitive and microeconomic modelling. To validate our approach, we consider some evidences from the real world and therefore decide to focus our analysis on three real life examples from Automobile manufacturing, Textile Industry and e-Marketplaces.

We concluded that inter-organizational networks are dynamic entities running into adaptation, where the internal synergies and the relationships with other networks shape the course of evolution. Networks are particular forms of organizations, in which firms jointly create both their destiny an the destiny of others and they come to see themselves as parts of business ecosystems.

Our main contribution focuses in the following aspects:

- Examples of literature on firm survival that relate *density dependence* with the development of relationships are few or inexistent. The complexity of analysing the setting up and evolution of networks is maybe one of the aspects that make this issue difficult;
- The bottom-up analysis, typical from Multi-Agent simulation is applied to a network-based approach in an economical context. This situation is original

in what concerns the use of networked Multi-Agent Systems (MAS) in an industry-based application;

- The cognitive perspective, with its several characteristics is unexplored in the literature of Organizational Ecology. This view offers the advantage of facilitating the individual modelling of the firm as well as a good way to understand firm behaviour.

The structure of this thesis is as follows:

In Chapter 2, we start with a general overview of the main concepts of firm and market (2.1), focusing on the organizational ecology perspective (2.2). Inter-organizational networks are described in Section 2.3.

Chapter 3 focuses on the links between *individual* and *aggregate* levels of analyses. In Section 3.1, we introduce the issue "individual versus social", where the individual is originally treated as an adaptive individual or organization. In Section 3.2., we describe the types of interactions between individuals, namely forms and methods of cooperation. In Sections 3.3, 3.4 and 3.5, our main goal is to provide a framework for the empirical applications. We introduce the simulation issues, where individuals are considered *simulated* agents and Multi-Agent Systems are provided.

In Chapter 4 we define the main questions and the modelling strategy of this work: we define the problem and the main questions that are considered as the research questions (4.1); an overview of the state of the art in the resolution of similar problems, is discussed in Section 4.2.; finally, in Section 4.3., the modelling strategy is introduced and discussed.

Two studies will be presented after that. In the first study (Chapter 5), a Cellular Automata-based application, CASOS, is used. Our main objective with this application is to analyse the effects of the contemporaneous and founding density on the mortality of organizations and determine the density survival limits of a particular niche.

In Chapter 6, we present the second study using a Multi-Agent System for modelling collaboration networks. Interactions between organizations are more complex, given that the framework includes a set of initial definitions, a microeconomic model, a decision-making process, and a cognitive model.

Chapter 7 contains conclusive remarks and directions for future work.

2. Organizations and Networks: an overview

In this chapter, we start with a general overview of the main concepts of *firm* and *market.* Then, the perspective of Organizational Ecology will be introduced and finally the main mechanisms that rule the creation of inter-organizational networks are described.

2.1. A general overview of the main concepts

A brief overview of the main organizational concepts that are studied in theory, namely *firms* and *markets,* is made in this section.

In a seminal work, with the aim of studying the function and nature of the firm, Penrose (1959) stated that "*the primary economic function of an industrial firm is to make use of productive resources for the purpose of supplying goods and services*".

Mainstream economics assumes that profit maximization is the goal of the firm. More recent analysis suggests that sales maximization or market share combined with satisfactory profits may be the main purpose of large industrial corporations.

The reasons for the existence of firms take into account the following aspects:

- internalization of risk;
- minimization of transaction costs;
- internalization of the benefits of knowledge;
- internalization of the benefits of size.

The control of determinant aspects as *risk*, *knowledge* and *size* is of crucial importance for firms. With respect to the minimization of transaction costs, a central factor for the existence of firms, Ronald Coase (Coase, 1937) represents one of the first attempts to define theoretically the firm in relation to the market. Within a firm, market transactions are eliminated and an entrepreneur, who controls production, substitutes the complicated market structure of exchange transactions. According to Coase (1937), a firm has a role in the economic system if transactions can be organized within the firm at less cost than if the same transactions were carried out through the market.

In the article "The logic of economic organization", Williamson (1991) analyses the work of Ronald Coase referred to above. Williamson's main contribution is the formalization and the systematization of several concepts that were presented by Coase. For instance, Williamson insisted on the need of identifying micro-analytic factors that are responsible for differences among transaction costs.

Within a firm perspective, these costs are internalized through a set of processes vertically integrated in a hierarchy. Discussion about markets and hierarchies are thoroughly analysed by several authors. Williamson has distinguished markets from hierarchies as two different and exclusive alternatives for economic organization[4].

[4] A short discussion about markets and hierarchies will be presented in the course of this work.

The essential difference between economic activity inside the *firm* and economic activity in the *market* is that "*the former is carried on within an administrative organization, while the later is not",* (Penrose, 1959). According to the perspective of organization studies, *Firms* are characterized by the fact that members (or actors) combine their resources and establish a corporate actor with collective decision-making rules, as well as rules for the allocation of income from the joint activities and resources (Ebers, 1997). *Markets* institutionalize competition among actors for opportunities of exchange.

However, there have been different perspectives in the study of these concepts. Mainstream economics focus on individual firm and attribute rational behaviour to firms in addition to profit motives. Organizational ecologists make no such assumptions. Organizations may compete successfully without maximizing any observable utility (unless the term 'utility' is tautologically defined to mean any consequence of organizational action). Ecological concern is most heavily concentrated on the probability of survival of the organizational form, not the individual organization.

In the following sections we analyse the perspective of organizational ecology where the fields of sociology, ecology and organizational theories were combined to provide a different description of the phenomena linking organizations and the environment.

2.2. The perspective of Organizational Ecology

In the following sections[5] we make a connection between different perspectives that have been fruitfully combined providing a new field called Organizational Ecology (OE). We introduce some important concepts such as selection, adaptation, isomorphism, competition theory, innovation, niche, generalism, specialism, inertia, legitimation and competition.

[5] In some of these section we are grateful for the preliminary dratf of Silva (2006).

Special attention will be given to the work of Stinchcombe (1965), Hannan and Freeman (1977, 1984), Freeman (1982, 1997), Carroll and Hannan (1989, 1992) and Wissen (2004), among other authors.

We will also focus on the aspects of *density dependence* because of its importance to one of the objectives of this work, namely the founding and survival of firms.

2.2.1. Social Structure and organizations

Organizational ecologists introduced a perspective in the analysis of organizations that is different from that of economists. Just as economists view the market as optimizing social utilities, organizational ecologists view competitive processes as optimizing fitness (Freeman, 1982).

According to Hannan and Freeman (1984), the ecology of organizations is an approach to the macro-sociology of organizations built on general ecological and evolutionary models of change in populations and communities of organizations. The goal of this perspective is to understand the forces that shape the structures of organizations over long time spans.

One fundamental step for exploiting the work in the fields of sociology, ecology and organizational theories was the contribution of Stinchcombe (1965), who has studied the relation of the society outside organizations to the internal dynamics of organizations.

For Stinchcombe (1965), the term *social structure*, means *"any variables which are stable characteristics of the society outside the organization"*, while, by *organization*, the author means "*a set of stable social relations deliberately created with the explicit intention of continuously accomplishing some specific goals or purposes*".

By analysing the society outside the organization, Stinchcombe introduced the idea of the environment where organizations live in, in the same way that ecology uses the *milieu* for the study of animals and plants. In this framework, Stinchcombe (1965) asks two main questions:

- How do Social conditions affect the degree of motivation that a population has to start new organizations?
- How do Social conditions affect the likelihood that a given foundation of an organization will survive?

For a summary on the social conditions affecting the founding of organizations, the author turns to general features of populations and social structure, for example: the level of literacy, the degree of urbanization, the existence of money economy, the incident of political revolution, and the level of past organizational experience in the population. Such characteristics affect motivation to start new organizations and the probability of survival of new organizational forms.

In general, a percentage of organizations that fail is higher amongst new ones. One of the most important aspects of Stinchcombe's work is the idea of the "Liability of Newness". In such cases, when this liability of newness is extremely large, organizations will tend to be accepted only under extreme conditions, for example in wartime.

New organizations involve new roles that have to be learned and the process of creating new roles has costs. Stinchcombe, subsequently, stressed the importance of standard social routines that reduce the liability of newness.

Besides that, as the author refers, the empirical evidence shows that certain structural characteristics of an organizational form are surprisingly stable over time. Where data are reasonably and readily available, structural characteristics of a type of organization tend to endure, and so there is a strong correlation between the age at which industries were developed and their structure at the present time. This is the concept of *organizational inertia* that will be discussed later.

2.2.2. Selection, Adaptation and levels of analysis

In a seminal paper written in 1977, Hannan and Freeman (1977) analysed the effects of the environment on organizational structure with a deeper look on ecological issues as a metaphor to study organizations. The ecological perspectives focused on *selection* while most of the literature on organizations adopted a different view, which the authors called the *adaptation* perspective.

Hannan and Freeman (1977) drew attention to the subtle relationship that exists between selection and adaptation by stating that *"adaptive learning for individuals usually consists of selection among behavioural responses"*.

Adaptation involves selection among certain possibilities. Almost all evolutionary theories in social science claim that social evolution is Lamarckian rather than Darwinian in the sense that human actors learn by experience and incorporate learning into their repertoires. "*To the extent that learning about the past helps future adaptation, social change is indeed Lamarckian – it transforms rather than selects. In other words, major change processes occur within behavioural units",* (Hannan and Freeman, 1984).

Combining and applying these two evolutionary perspectives, Usher and Evans (1996) demonstrated how Darwinian processes at the unit level, may lead to Lamarckian adaptations at the organization level, through purposeful replication of successful forms. The issue here is to find a balance between the relative contributions of two fundamental processes that work towards the reconfiguration of populations of organizations: Selection/replacement and transformation.

The scope of this empirical study from Usher and Evans, included the following types of organizations: (i) gas stations in Edmonton, Alberta, from which gasoline was available for retail sale at any time between 1959 and 1988; (ii) new forms of gas stations that arose meanwhile; (iii) changes that took place in individual stations.

Whether changes are intentional or accidental, managers viewed the outcomes of these changes and selectively retained those that they believed would be the most beneficial, thus effecting changes in individual organizations. A comparison between Lamarckian and Darwinian processes would conclude that:

- *"via the Lamarckian process, an organizational population changes as existing members adapt to environmental pressures by replacing less favoured competences with more favoured competences".*
- *"via the Darwinian process, an organizational population changes as new members holding more favoured competences compete into failure members holding less favoured competences"*

According to the adaptation perspective, subunits of the organization, typically managers, or dominant coalitions, examine the relevant environment, for opportunities and threats. Then, they define strategies and adjust organizational structure. Hannan and Freeman (1977) clearly recognize that leaders of organizations do prepare strategies and organizations to adjust to the environment.

Therefore, at least some of the connection between structure and environment must reflect adaptive behaviour or learning. However, there is no reason to presume that the vast structural variability among organizations reveals only or even mainly adaptation. To support this, Hannan and Freeman (1977) claim the existence of several obvious limitations on the ability of organizations to adapt.

Levels of analysis in OE

Ecological analysis is introduced at three levels – individual, population and community. Events at one level almost always have consequences at other levels; however it is impossible to reduce population events to individual events, as individuals do not reflect the complete genetic variability of the population. The study of organizations is more complex since the research faces at least five levels of analysis:

- *members,*
- *subunits,*
- *individual organizations,*
- *populations of organizations*
- *communities of (populations) organizations.*

The extra complexity occurs because organizations tend to be more easily decomposable into component parts than organisms. Also, individual members and subunits may move from organization to organization in a way that has no parallel in nonhuman organizations.

Some of these concepts are not easy to define. In fact, Hannan and Freeman (1977) admit the difficulty in defining "*population of organizations*". The authors recall that all organizations are distinct and so no two are affected identically by any given exogenous shock. Even so, they consider the identification of relatively homogeneous classes of organizations (shall we call them "representative ones") in terms of environmental susceptibility as viable.

According to the authors, it is important for the study of organizational populations to make an analogy to the biologist's notion of species. Monod (1971)'s suggestion is that the genetic content of any species is a blueprint which holds the rules for transforming energy into structures. Therefore, the adaptative capacity of a species is summarized in the blueprint. Based on Monod's insight, organizational theorists argue that to identify a species analogue for organizations, it is necessary to search for such blueprints, that is, the rules or procedures for obtaining and acting upon inputs in order to produce an organizational product or response.

The way how this blueprint can be inferred within organizations is usually done by examining the formal structure of the organization in the narrow sense, that is: (i) its tables of organization, written rules of operation, etc; (ii) the guides of activity inside the organization; (iii) the normative order, i.e., the ways of organizing that are defined as right and proper by both members and relevant sectors of the environment.

A complete analogy to species demands a search for qualitative differences among the formal structure. Hannan and Freeman (1977) consider that such differences mainly appear in the formal structure and the normative order that feature the organization. In the next paragraphs, we give a more detailed analysis of this formal structure, considering different *forms* of organizations.

2.2.3. Isomorphism , Competition theory and Innovation

One of the concepts explored by the Organizational Ecology is the '*Form*'. But what exactly is a form? Although the literature uses the form notion indiscriminately, little attention has been paid to clarify and develop this theoretical concept. Following Carroll and Hannan (2000), "*forms are abstract specifications of types of organizations. Populations are concrete manifestations of the types, bounded in time and place"*.

A complete analogy to species is possible within the context of Organizational Ecology but demands a search for qualitative differences among forms. Hannan and Freeman (1977) consider that such differences mainly appear in the formal structure and the normative order that characterizes the organization. Adaptation is a concept that may be measured in terms of *fitness*. Fitness is then defined as the *probability that a given form of organization would endure in a certain environment.*

In what concerns new forms of organizations, Freeman (1982) stressed the importance of the distinction between two different types of organizational birth: *creating new forms and copying existing forms*. Freeman's goal is to analyse the role of organizational innovation, and therefore he has defined "life cycles processes" and "natural selection processes" as follows:

- **Life cycle process**: the patterns over time through which new organizations come into being, change, and disappear.
- **Natural selection**: the differential reproduction and survival of organizations depending on relative competitive advantages.

This distinction leads to the concept of entrepreneurship and the employment relationship. Entrepreneurs rarely begin without extensive experience in similar kinds of organizational activity.
Freeman also studied *innovation* as the development of new forms and described several models of growth and development of organizations. "*The importance of such models for the purpose at hand is that they provide a rationale for differences among organizations* (…). *Innovation is the development of new forms"*

Selection may operate in such way that the proportion of organizations affected increases with population size. This is called *density-dependent* selection'. Therefore, if organizations change according to the number of already existing organizations, then the form may also change as organizations grow. Freeman concludes: *"We expect to find diversity of organizational forms, then, either because growth leads to change of form (metamorphic development) or because single organizations have difficulty growing fast enough and simultaneously being efficient enough to prevent the appearance of competing forms of organizations"*.

Another concept studied by Hannan and Freeman (1977) is *isomorphism*, which constitutes an answer to the question: "*why there are so many kinds of organizations?"* In fact, the diversity of organizational forms is isomorphic to the diversity of environments. This means that each unit experiences constraints which force it to resemble other units with the same set of constraints. Thus, isomorphism is a good concept only in stable environments.

Hannan and Freeman (1977) propose a formulation of isomorphism based on the mechanism or mechanisms responsible for equilibrium. In this respect, the principle of isomorphism must be supplemented by a criterion of selection and *competition theory***.**

In what concerns competition theory, Hannan and Freeman (1977) recall the formal framework of Lotka-Volterra equations. They evoke two distinctive ecological considerations: '*the capacity of the environment to support forms of organization and the rate at which the populations grow (or decline) when the environmental support changes*.' To explain these considerations, Hannan and Freeman start with the models of population dynamics that were independently proposed by Alfred Lotka (Lotka, 1925) and Vito Volterra (Volterra, 1927) that incorporate the effects of competition between populations. Assuming that population growth of isolated populations has a S-shaped growth path, the growth rate is given by the product of a growth rate and the current size of the population:

$$\frac{dN}{dt} = rN \qquad \textbf{Equation 2-1}$$

where N denotes the size of the population and r is defined as the difference between the birth and the death rates of the population. This model is the basis for other population models, as the logistic model of population growth, that incorporates the carrying capacity, K[6]:

$$\frac{dN}{dt} = rN\left(\frac{K - N}{K}\right) \qquad \textbf{Equation 2-2}$$

This parameterization provides a substantial appealing to introduce competition: two populations compete if the size of each population lowers the carrying capacity of the other. Therefore, a new formulation of the Lotka-Volterra model can be considered for competitive interactions. It assumes that the effect of the density of the competitor on the realized carrying capacity is linear. In the case of two populations we have:

[6] The carrying capacity, K is the maximum size a population can attain, under the conditions of the current environment.

$$\frac{dN_1}{dt} = r_1 N_1 \left(\frac{K_1 - N_1 - \alpha_{12} N_2}{N_1} \right) \quad \textbf{Equation 2-3}$$

$$\frac{dN_2}{dt} = r_2 N_2 \left(\frac{K_2 - N_2 - \alpha_{21} N_1}{N_2} \right) \quad \textbf{Equation 2-4}$$

in which N_1 and N_2 are the sizes of the two populations, K_1 and K_2 represent the corresponding carrying capacity of the two populations, and α_{12} and α_{21} are the competition coefficients for competing populations that represent how increases in one population affect the growth of the other.

A stable two-population equilibrium exists for the system if,

$$\frac{1}{\alpha_{21}} < \frac{K_2}{K_1} < \alpha_{12}. \quad \textbf{Equation 2-5}$$

Hannan and Freeman (1977) extend the Lotka-Volterra system to include M competitors and develop a model that demonstrates that when growth in population is limited only by resource accessibility, the number of distinct resources sets an upper bound on diversity in the system.

2.2.4. Niche theory, generalism, specialism

Bridges between ecological theories and organizations have been built and researchers tried to identify important features that allow them to define and characterize organizational populations. One of these features is *organizational niche*. In Ecology, a niche describes the relational position of a species or population in an ecosystem, including how a population responds to the abundance of its resources and enemies, etc. In organizational studies, a niche is characterized by a set of organizational capabilities and a location in a resource space (Baum and Singh, 1994). It includes all the combinations of resource levels at which the population can survive and reproduce itself. The location is important and each population occupies a different niche.

Niche theory is also related with the dynamic of the species. Following Hannan and Freeman (1977), in terms of the population characteristics over time, two kinds of phenomena can occur: generalism and specialism. In the generalism view we can observe a wider band in environmental variation, while specialism implies a smaller variation.

Hannan and Freeman (1977) showed interest in determining which points will be favoured by natural selection and made some effort to model the optimization process they distinguish two situations, in terms of environment:

- **fine-grained**: when an organization encounters many units or replications; from a temporal perspective, we can say that typical durations in states are short relative to the lifetime of organizations; it is better for specialisation.
- **coarse-grained**: when the duration of environmental state changes is long, generalists need not spend most of their time and energy altering structure. It is better for generalisation. Coarse-grained and uncertain variation favours a distinct form of generalism: polymorphism.

As we have stated before, *location* is also an important feature for the definition of an organizational population. Organization environments have spatial components that affect the evolutionary dynamics of organizational populations. Lomi (1995) investigated the effects of location dependence on founding rates of rural cooperative banks in Italy from 1964 to 1988.

Baum and Mezias (1992) found that hotel failure rates have increased significantly and this was due to the effects of localized competition in terms of geographic location, size, and price. Localized competition was measured by comparing the position of a focal hotel to the position of the others within a given distance. They found that more similarly sized, priced, and located hotels compete more intensely.

We will come back to this subject later in this work and will take a closer look at the relation between niche theory and the density dependence theory.

2.2.5. Inertia and Change

Studies have been developed where the growth and survival of organizations were the main research topics. Freeman (1982) studied the evolution of organizations and described the organizational life cycles and natural selection processes.

Hannan and Freeman (1984) introduced the concept of *structural inertia,* which is defined in the following way: inertia, like fitness, refers to a correspondence between the behavioural capabilities of a class of organizations and their environment. In fact, *"structures of organizations have high inertia when the speed of reorganization is much lower than the rate at which environmental conditions change*".

In their article Hannan and Freeman search for an explanation for the structural inertia due to the organizational change. The major assumption of this idea is that individual organizations are subject to strong inertial forces, that is, they seldomly succeed in making radical changes in strategy and structure in the face of environmental threats.

The authors argue that selection processes tend to favour organizations whose structures are difficult to change. In addition, they state that organizations with high levels of structural inertia are an outcome of an ecological-evolutionary process. Hannan and Freeman (1984) also studied the effects of age, size and complexity in structural inertia. Based on several assumptions for selection and reproduction, they have proposed the following theorems:

- ***Theorem 1****: Selection within populations of organizations in modern societies favours organizations whose structures have high inertia.*
- ***Theorem 2****: Structural inertia increases monotonically with age.*
- ***Theorem 3****: Organizational death rates decrease with age.*

- ***Theorem 4****: Attempts at reorganization increase death rates.*

The third theorem is often called the "liability of newness" – a very well recognized hypothesis proposed by Stinchcombe (Stinchcombe, 1965).

Size and complexity also play an important role in organizational inertia. Hannan and Freeman (1984) assumed that small organizations are more likely to die when they start the process of reorganization. On the other hand, "complexity" (defined as patterns of links among subunits) increases the risk of death. The level of complexity is proportional to the number of links - delimiting a network - among the members of the community: firms or subunits inside an organization.

2.2.6. Legitimation and Competition

Legitimation plays an important role in theories of Organizational Ecology. Following Wissen (2004), *legitimation* refers to the degree that a new organizational form is known and accepted in society[7]. In spite of this quite simple definition, legitimation and competition are two concepts that researchers find difficult to measure. Carroll and Hannan (2000), distinguish two types of legitimation:

- *an organizational form can receive legitimation to the extent that its structure and routines follow the prevailing institutional rules (coercive isomorphism)*
- *an organizational form gains legitimation when it attains a social taken-for-granted character (constitutive legitimation).*

[7] According to Poston and Micklin (2005) the concept of Legitimation can be explained as follows: *when a new industry emerges, customers are not familiar with the product, investors are reluctant, and there may be legal or institutional constraints that prevent free market introduction. Legitimation increases with the number of firms: the product becomes more familiar for customers, knowledge increases and investors become less reluctant. Founding and disbanding are related to the level of legitimation of an organizational form.*

These two concepts differ, but authors say that the conception of constitutive legitimation has been more interesting for research in organizational demography because it has greater connection to density.

For example, when automobile manufacturers appeared in the nineteenth century they usually lacked constitutive legitimation and this usually makes organizing more difficult. Following Carroll and Hannan (2000), "*capital sources are wary, suppliers and costumers need to be educated, employees might be hard to identify and recruit; and in many cases hostile institutional rules existed"*. Legitimation increases as the new form is disseminated. The way as legitimation (and competition) relates with the founding and death rate is explained in the following sections.

Competition refers to some kind of negative effects of the presence of one or more actors on the life chances or growth rates of some local actor. In the context of demography of organizations, there are two kinds of competition:

- *Structured or directed*, when two actors engage in rivalry or head-to-head competition;
- *Diffused or undirected, when a set of actors are dependent upon a pool of limited resources.*

Both structured and diffused competition are important to understand the life chances of organizations but the demographic theory of density dependence focuses on diffused competition because it arises as a function of the number of potential bilateral competitors. So, the concept of competition will mean, in general, diffused competition.

There is clearly a positive relationship between the number of firms and the level of competition. As the size of a population increases linearly, the level of competition may grow geometrically. This means that adding a new firm to a large population has more impact than adding a new firm to a small population (Wissen, 2004).

In the next section the relation between competition and vital rates (birth rate, death rate) will be presented. In other words, we will take a closer look in the density dependence process, which is vital to the analysis of the survival of the organizations.

2.3. Density Dependence Processes

In this section, we analyse the importance of the density dependence process on the survival of the organizations. First, we define the concept of *Density Dependence* and then relate it with to the concepts of legitimation and competition that were introduced in the previous sections.

Different perspectives are studied: Wissen (2004) introduced a spatial interpretation of the density dependence that relates it to economies of agglomeration in industrial economics. Other authors (as Baum and Singh, 1994), analyse the importance of density in the context of organizational niches.

2.3.1. The importance of density dependence

Organizational density of a population is defined as the number of organizations in the organizational population. Density dependence is introduced in the study because of its importance to the founding and failure of firms. In this section, we start by stressing the role of density in organizational demography and then relate this concept to those of founding, failure, legitimation and competition.

In human demography there is a concept named *demographic transition* which refers to a specific chronological sequence of changes in vital rates (death rate, birth rate, fertility rate). In pre-industrial stages, mortality rate (number of deaths in a population of given size) and fertility rate (number of births relative to a given population of women) are high, so there is a great amount of deaths and births.

The demographic transition consists of a decline of the mortality rate, followed by a decline in the fertility rate. As long as societies undergo this transition, the rate of population growth changes from high to low. There are some details in this transition that depend on the vital rates and the age structure of the population (Keyfitz, 1977), but what is interesting in this phenomenon is that it produces rapid transitions in the population growth because of particular patterns of historical time dependence in vital rates.

In corporate demography researchers have found that similar historical periods of rapid organizational population growth often arise because of particular patterns of density dependence in vital rates (Carroll and Hannan, 2000). One of the main efforts has been to come up with the general model of a long term organizational evolution: the density model of legitimation and competition.

As Freeman (1982) puts it, "*Density dependence is important because it generates homeostatic processes in populations, that is, it generates equilibrium levels toward which population sizes adjust, usually at decelerating rates*". Density dependence can be defined as follows:

$$\text{Density dependence} = \frac{dN}{dt} = rN\left(\frac{K-N}{K}\right). \qquad \textbf{Equation 2-6}$$

As we saw before, this model is associated with the S-shaped logistic growth curve that is used to model situations in which the growth of the population is limited to K (the carrying capacity, that is to say, the maximum size a population can attain, under the conditions of the current environment). N represents the population size and r is the intrinsic rate of population growth.

Equation 2-6 (named Verhulst equation) constitutes a typical application of the so called logistic equation, often used to model the growth of population . It states that the rate of reproduction, r, is proportional to the existing population, N, and that the rate of reproduction is proportional to the amount of available resources, all else being equal.

Thus the second term models the competition for available resources, which tends to limit the population growth. The solution of this equation is proved to be K (Hannan and Freeman, 1989).

According to Carroll and Hannan (2000), this model is deficient in a least three ways:

- The model does not consider the effects of competing organizational forms (on the contrary, the Lotka Volterra model captured by equations 2.3. and 2.4 does);
- The model does not take into account variations in carrying capacities (i.e., the limits of the population according to the limited stock presented in equation 2.1);
- The model does not consider organizational growth or decline.

2.3.2. Relation with legitimation and competition

Carroll and Hannan (1989) presented formal models of density dependence using legitimacy (L) and competition (C) as explanatory variables. In particular, they argued that the founding rate is directly proportional to legitimacy and inversely proportional to competition. The reason for that to happen with legitimation is that a taken-for-granted form can be more readily visualized by potential organizers than one with unknown reputation. Therefore, more organizations will be born when legitimation grows.

In what concerns diffused competition, organizational analysts believe that intense competition causes supplies and resources to become exhausted and markets get packed tightly (Carroll and Hannan, 2000). So, the founding rate in an organizational population is inversely proportional to the intensity of diffuse competition. We also know that competition increases with density at an increasing rate.

In equation 2.6., $\lambda(t)$ denotes the founding rate at time t and $\alpha(t)$ is a function that summarizes time-varying environmental conditions. L_t designates the legitimaton and C_t, the competition.

$$\lambda(t) = \alpha(t)\frac{L_t}{C_t}$$ **Equation 2-7**

Following the argument that organizational density (n) increases legitimacy at a decreasing rate and increases competition at an increasing rate, Hannan (1986) proposed the following parametric relationships:

$$L_t = n^{\beta}_t$$ **Equation 2-8**

$$C_t = \exp\ (\gamma n^2_t)$$ **Equation 2-9**

where β and γ are constant parameters, with $0<\beta<1$ and $\gamma>0$. Combining these two equations, we can obtain a model for the effect of density in founding rate:

$$\lambda(t) = \alpha(t)\ n^{\beta}_t \exp\ (-\gamma n^2_t)$$ **Equation 2-10**

with $\gamma<0$ and $0<\beta<1$.

According to this model, the founding rate has a monotonic inverted U-shape relationship with density: initial growth in density increases the founding rate but further increases eventually depress it.

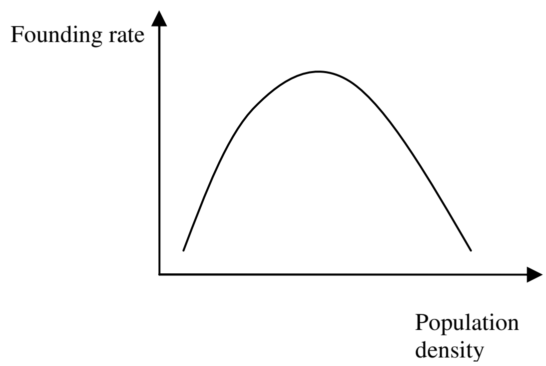


Figure 2-1 – Relationship between the founding rate and population density

Similar arguments hold for the mortality rate – but with the rate proportional to C and inversely proportional to L. Carroll and Hannan (1989) revised a previous model from Hannan and Freeman (1988) with the following form:

$$\mu(a)= \zeta(a)\ \exp(\theta_1 n_a+\theta_2 n^2_a) \qquad \textbf{Equation 2-11}$$

where θ_1 and θ_2 are constants, with $\theta_1<0$ and $\theta_2>0$; $|\theta_1| > |\theta_2|$, n_a denotes the density of the population at time the organization attains age **a** and $\zeta(a)$ summarizes the effects of of environmental conditions operating at that time.

The mortality rate, $\mu(a)$, has a monotonic U-shape relationship with density as shown in Figure 2-2.

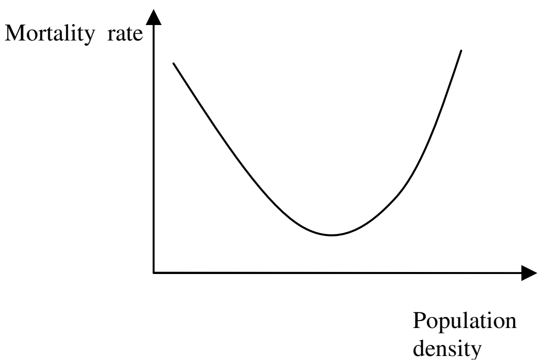


Figure 2-2 – Relationship between the mortality rate and population density

Actually, these models of founding and failure belong to a more general family of density dependence models with the form:

$$\Psi_t = \exp\ (\alpha n_t+\beta n^2_t). \qquad \textbf{Equation 2-12}$$

When Ψ_t is the rate of entry or growth, the model predicts that $\alpha>0$ and $\beta<0$. When Ψ_t is the rate of failure or exit, the model predicts that $\alpha<0$ and $\beta>0$. A possible interpretation of the two parameters α and β is that they represent, respectively, the effects of the *legitimation* (the institutionalization of an organization form) and *competition.*

These relationships have been well explained by Wissen (2004) that argues that these two basic forces (legitimation and competition) are responsible for the dependency between the population density and founding (birth) and failure (death) of firms. The founding rate is directly proportional to the level of legitimation whereas the mortality rate is inversely proportional to the level of legitimation; opposite associations are observed with respect to competition.

Furthermore, the joint effects of legitimation and competition explain to a large degree the shape of the density function over time. As the population grows, legitimation increases and competition remains very restricted, and so the growth rate increases (see dotted line in Figure 2-3). At a certain level (1), the maximum level of legitimation is reached, and competition starts to increase fast. Consequently, the growth rate decreases quickly to zero or even to a negative value (2), and the density falls down to very low values. Population density function has a specific S-shaped, attained before the level (1).

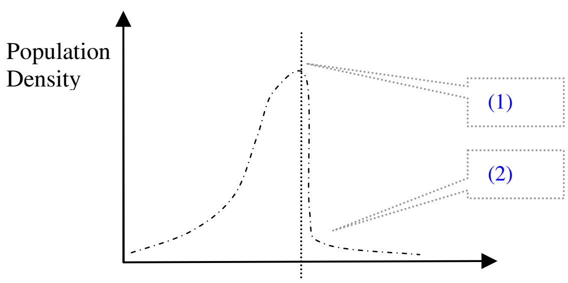


Figure 2-3 - S-shaped form of density over time (until instant t'), followed by a fast decrease to near zero

The point (1) corresponding to the maximum level of the density is named the *carrying capacity.*

2.3.3. Delayed density dependence

Carroll and Hannan (1989)'s main concern was whether density at the time of founding might change the mortality rate of adult organizations. The authors considered some mathematical specifications of the mortality rate (Gompertz and Weibull models) to analyse empirically three types of industries. The authors have called this phenomenon of postponement between the cause (density at founding) and the effect (contemporary death), *delay density model*.

The most important variables of this density delay model are the density at the time of founding (*delayed density*), density at present time (*contemporaneous density*) and *organization age*. They have concluded that density at time of founding had a positive effect on mortality rates in all the populations that have been studied.

Some years later, Lomi and Larsen (1998) introduced another approach regarding the delayed density dependence process using a model of cellular automata[8]. Synthetic data was produced according to a process of life and death in a grid where an organization lived if the number of organizations around it belonged to a certain interval (the same applied to new births).

Data was applied to the Gompertz and Weibull model and some parameters have been estimated. In spite of some lack of agreement with reality of this cellular automata model, authors argued that the local interaction represents a good means to understand the phenomenon of density processes. Furthermore, Lomi and Larsen (1998) claimed that the connection between density at founding and mortality rates is an emergent feature of the evolution of organization populations. This model will be brought to analysis again in Chapter 4.

---

[8] As we will see later in Chapter 3, *Cellular Automata* are cells located in a regular grid where the behaviour of an individual cell is determined by a set of rules which specify how that state depends on the previous state of that cell and the states of the neighbour cells.

2.3.4. Introducing other variables in the models

Density plays an important role in Organizational Ecology but it is not always the main determinant of firm survival. Bearing on the key question of how the number of firms in a population and the population density affect the intensity of competition within the industry, Barron (2001) compared three dynamic models of organizational populations. Some of the models used *size* and *age* of firms too. Monte-Carlo micro-simulations have been used and the vital rates (founding, failure, and growth/decline) have been modelled as stochastic processes taking a regression model for each vital rate.

The corresponding hazard rate function, $\mu(a)$, depends on the population density, age and size of the organization, and other variables required by the three models. In the models compared by Barron, Ψ_t represents the rates of entry, growth or exit, depending on the signal of the involved parameters and N_t represents[9] density at time t:

Barnett's model: $\Psi_t = \exp(\alpha N_t + \beta N_t^2 + \gamma_1 M_t + \gamma_2 T_t + \gamma_3 Q_{it})$ **Equation 2-13**

Hannan's model: $\Psi_t = \exp(\alpha N_t + \beta N_t^2 + \theta_1 N_t \times t + \theta_2 N_t^2 \times t)$ **Equation 2-14**

Barron's model: $\Psi_t = \exp\left[\alpha N_t + \beta N_t^2 + \xi \log(S_{it}) + \pi N_t \times \log(S_{it})\right]$ **Equation 2-15**

In these specifications, $T_t = \sum_{i=1}^{N_t} A_{it}$ is the sum of organization ages (being A_{it} the age of the firm i at instant t) ; $M_t = \sum_{i=1}^{N_t} S_{it}$, is the population mass, i.e., the aggregate size of all the organizations in the population (S_{it} being the size of firm i at instant t); and $Q_{it} = \log\left[\sum_{i=1}^{N_t}\sum_{j=1}^{N_t} A_{jt} S_{jt}\right]$.

As it can be seen, some of the models do not use all the variables. In the Barnetts' model, as well as in the Barron's model, the effect of the size appears as an independent

[9] The notation N_t instead of n_t is often used to represent density (total number of organizations in a given area) at time t.

variable. The variable *age* is considered only in the Barnetts' model, although we may consider the age as an implicit variable in these models, because as time goes by the firms get more mature.

The parameters of these models have been estimated using real data through maximum likelihood techniques (Barron, 2001). Barron concludes that all three models work well when looking at the founding rates (some different values were obtained for death rates), but it is difficult to distinguish the models in terms of a statistical test.

2.3.5. Critics to the density model and the spatial interpretation

The density dependence model has been criticised, although recently some other variables were introduced. Wissen (2004) puts forward a criticism of the model by saying that: "*Legitimation and competition explain the S-shaped form of population growth which leads to a stable population size at the level of the carrying capacity, but no explanation is given to the decline of the growth rates in the peak of density curve since a decrease in the population size would lead to less competition and therefore a return of the growth rate to zero*."

Furthermore, *size* is not taken into account in the theory, whereas clearly large and small firms have very different chances in a population. Firms differ not only with respect to size and economic activity, but also with respect to geographical location, which may be labelled spatial heterogeneity. Lomi and Larsen introduced spatial proximity into ecological models of populations, but they do not take into account agglomeration economies explicitly, which can be viewed as an extension of density model at the local level.

In the next paragraphs we will introduce the concept of *agglomeration economies* and then we will present the niche theory. In both cases, analysts consider the geographical proximity of firms.

Wissen (2004) compared the density dependence model from Organizational Economics (OE) with the notion of agglomeration economics which deals with the aggregation or concentration of firms in an economy. Both (density dependence model and agglomeration economies) share some features because both involve some form of positive feedback between size of the population and growth.

In Economics, there are two important concepts associated with agglomeration or aggregation economies: *localization economies* and *urbanization economies*. Localization economies result from the geographical clustering of similar types of firms, while urbanization economies involve firms from different industries (regional diversification). In urbanization economies, inter-firm interaction is very important in terms of input-output relationship of suppliers and deliverers.

Wissen (2004) compared density and location theories from industrial economics and tried to discover in what way one theory can improve the other. As we saw before, legitimation (which is not an economical concept) may have two meanings in OE: *conforming to a set of rules or conventions* and refer to *constitutive legitimation* (taken-for-grantedness). The author has identified these indicators of legitimation that resemble in many aspects the emergence of agglomeration economies.

Legitimation also involves the establishment of networks of other firms not only in input-output market relations, but also in various networks, involving entrepreneurs and organizations. In the next figure, the differences involving legitimation and agglomeration approaches are apparent (see Figure 2-4):

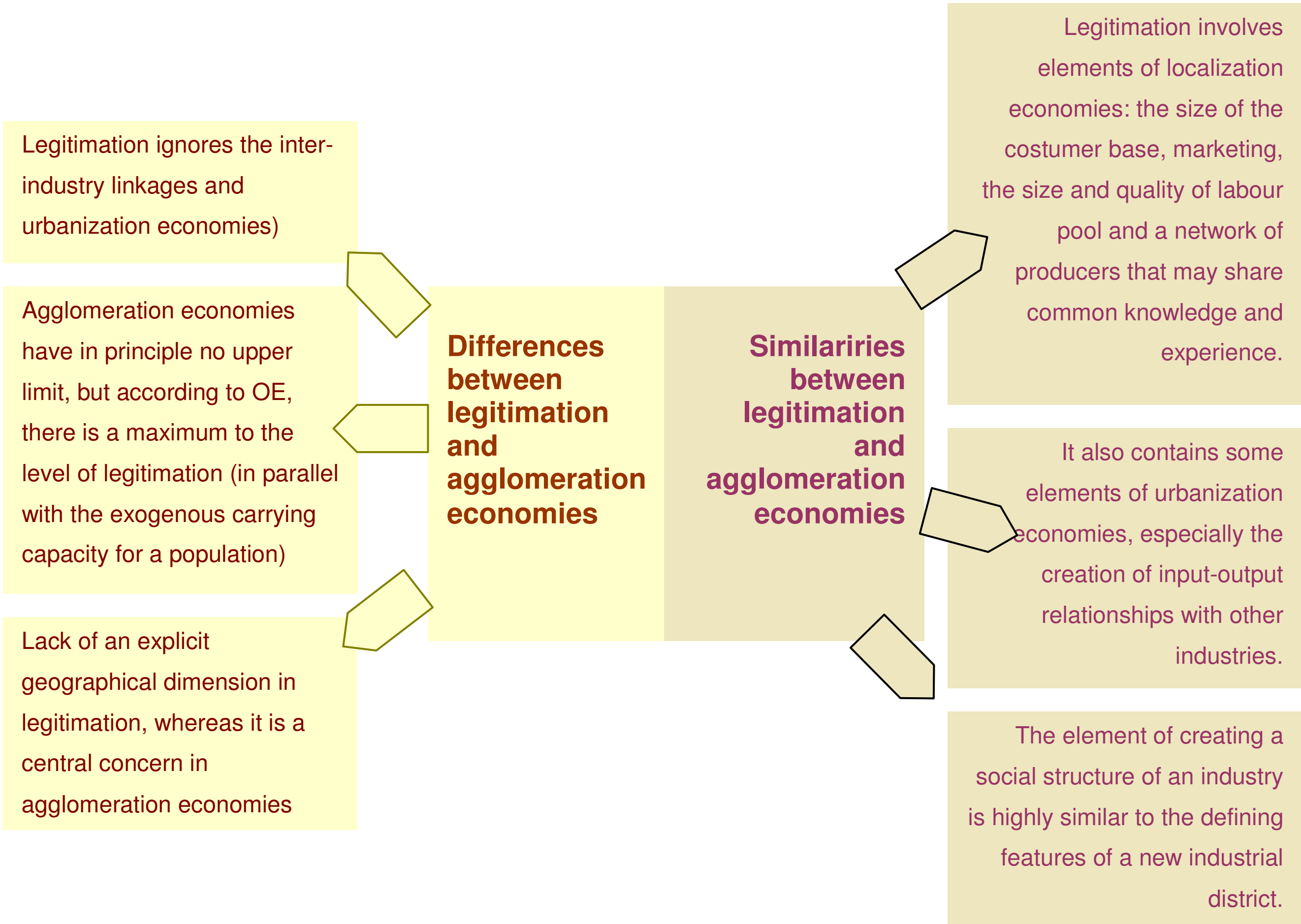


Figure 2-4 – Main differences between legitimation and agglomeration economies.

We could summarize the aspects about agglomeration effects in OE and density effects in models of agglomeration economies, as follows:

- In OE, the spatial dimension has not been very important and attention has been solely devoted to temporal variation in population change.
- Agglomeration effects are stronger in knowledge intensive industries.
- There are examples of spatial analysis within OE: a related development views resources as heterogeneous and deals with market partitioning into segments; for instance, we note the resource partitioning model (Carroll and Hannan, 2000).

Finally, to introduce the notion of agglomeration economies into the density dependence model, a number of extensions are necessary:

- Processes of legitimation and competition must be defined geographically.

- Interactions between firm populations should be incorporated in order to capture urbanization economies.
- Local variations in population density can also have a positive effect on founding and firm growth, especially in the early life cycles of the industry and in knowledge intensive industries.

In the following paragraphs, we provide a more detailed analysis of organizational niches and their importance to density, competition and legitimation.

2.3.6. Organizational Niches and density

Usually firms occupy niches in the market, which may be defined as geographical niches. In the empirical analysis of geographical heterogeneous populations two strands of research can be observed:

- A statistical approach, where the optimal size of the geographical region is determined as a result of an analytical model, as in Lomi (1995).
- An approach where, based on a priori examination, the size of the market area is fixed (Baum and Mezias, 1992).

Baum and Singh (1994) studied how organizational niches within populations influence patterns of competition and mutualism. A population of day care centres was analysed and a density model was implemented and tested. The density dependence model focuses on the dynamics of competition (organizations diminishing each other's fate) and mutualism (organizations enhancing each other's fate) to describe the evolution of organizational populations.

Several authors (e.g. Hannan and Freeman 1989, Hannan and Carroll, 1992) proposed that initial increases in density produce mutualism by increasing the institutional legitimacy (consitutive legitimation) of a population. As the number of organizations grows, competition for scarce common resources begins. Baum and Singh (1994) refer the effect of geographical proximity too, but the focus of their research is organizational niche, which is characterized by *a set of organizational capabilities and a location in a resource space*.

Authors refer the problem of overlapping organizational niches, as a way to capture competition. Overlap density is thus defined by the equation:

$$\text{Overlap density} = N_{it} + \sum_{j \neq i} w_{ij} N_{jt} \,, \qquad \textbf{Equation 2-16}$$

where N_{it} is the number of organizations in organizational niche i at time t, N_{jt} is the number of organizations in organizational niche j at time t and w_{ij} is the weight of organizational niche overlap of organizational niche i with organizational niche j. Overlap density depends on the organizational niche because competition occurs within the same niche and differs among competitive landscapes. Overlap density is also time variant.

The authors establish the hypothesis that "*overlap density is positively related to the mortality rate*", because greater overlap density implies greater competition for resources between a specific organization and all the other organizations in the population, and so competition for resources can influence the likelihood of mortality. In fact, organizations that attract more resources are more efficient at converting these resources into high quality services than others.

Baum and Singh (1994) have used a hazard function[10] to analyse the instant rate of failure. At the population level, the results of this study support the idea that there exists a curvilinear relationship between population density and failure rates. Increases in the population with differential organizational niches can lead to the emergence of an organizational system comprising organizations that work together because they are complementary.

---

[10] Hazard functions measure the failure rate and will be introduced and estimated later (in Chapters 5 and 6) in order to analyse the impact of some variables on the mortality of firms.

2.4. Inter-organizational networks

Introduction

In organizational ecology (OE), populations of organizations are studied and information about their evolution is of interest. In this section we introduce inter-organizational networks, a form of interaction among firms through the definition of inter-organizational relations. The network concept is a powerful metaphor, permitting to capture the connection between entities in space (Purchase and Olaru, 2003).

The *liability of newness* (Stinchcombe, 1965) and the *liability of smallness* (Baum 1996) concepts (some given in the previous sections), assume that a new firm does not have sufficient resources. The strategic use of external resources through inter-organizational networks provides an important base for growth. For that reason, the survival and growth of an organization depends on its capacity to maintain and extend its network of inter-firm relationships (Venkatamaran and Van de Ven, 1998).

Networks have been extensively studied from different perspectives: graph theory, transport economics, social sciences, etc.. Earlier, Kamann and Strijker (1991) classified the research on networks into four domains: regional science, infrastructure and logistics, industrial marketing and social science. However, nowadays the application of the network metaphor is observed in almost all disciplines and sciences.

In the past, division of labour in agriculture and in industrial society can be seen as a form or networking and a prerequisite for increasing prosperity. As stated by Smith (1776), the division of labour is limited by the geographical proximity of the market. Nowadays, the introduction of transportation technologies and information technology are driving the physically disintegrated markets and enterprises to global markets enabling thus maximal specialization of enterprises (Osterle *et al*. 2001)

In this section, we begin with the definition of the concepts related to inter-organizational networks, structured according to different perspectives: the first one is the system of relationships perspective based on the work of authors such as Hakansson, Gulati and Nohria, and on Williamson's transaction costs theory.

After that, we introduce the idea of *collaboration* as a fundamental concept in networks. Collaboration is associated with innovation and technological development as a result of exchange among different agents with the goal of overcoming the problems of a company related to the loss of competitive advantages. Finally, the concepts related to the network structure and its topology are described and analysed.

Motivation and Concepts

There are several approaches that can be pursued when studying inter-organizational networks: we give an overview of the theory and systematize the different points of view on this area concerning the issues of industrial economics to management[11].

In this section, we introduce the concept of *network* as it is analysed by organizational theorists such as Mitchel (1969), Hakansson (1982, 1987), Hakansson and Snehota (1995), Hakansson *et al*. (1999), and Koleva (2002), that recognize social networks as the basis to organizational networks. The section is divided into several parts, according to different perspectives that are followed.

[11] The IMP Group has given good contributions in the study of inter-organizational networks. IMP Group (www.imp.org) is an informal international group of scholars concerned with developing concepts in the field of business-to-business (B2B) marketing. More recently, the IMP group incorporated the study of complex networks within which business communities operate.

2.4.1. The concept of inter-organizational network

An inter-organizational network is a set of firms that interact through inter-organizational relations (Eiriz, 2004). Johanson and Matsson (1987) describe the network as a system of relationships based on a division of work in the network.

The notion of inter-organizational network is applied to a wide variety of relationships among organizations. The concept of inter-organizational networks can be applied to joint ventures, strategic alliances, industrial districts, consortia, social networks and others. In summary, it can be used to define any set of recurring ties (resource, friendship and informational ties) among a set of nodes (individual, groups, organizations, information systems, etc.).

Hakansson (1982, 1987), Hakansson *et al.* (1999) has studied the importance of relationships and learning in networks, giving important contributions to the study of inter-organizational networks. According to Hakansson (1987), a network of organizations is made up of three classes of basic variables, namely: *agents* (that manage activities or control resources), *resources* and *activities*[12].

Agents can be individuals, groups or organizations. It is assumed that the generic goal of an agent is to control other agents, allowing the mobility of resources for specific purposes. According to this view, activities permit agents to combine, develop or exchange other resources. Agents divide activities among them and must complete them in the most efficient way. Resources are prior conditions to the accomplishment of the activities. Resources are heterogeneous and there are many possibilities of how these can be combined. There are tangible resources (products, equipment) and intangible resources (knowledge, experience, contacts).

Kamann and Strijker (1991) give examples of two categories of agents (called *actors*) in a network: *market actors* (individual entrepreneurs, firms and subsidiaries, profit centers, or business units belonging to a corporate network) and *non-market-actors*

[12] The first letters of each word give rise to the acronym - ARA model.

(journalists, politicians, lobbies). These actors operate on a plain and abstract economic space, where they interact with other actors.

Based on Hakansson's approach and some empirical studies about networks, Ratti (1991) states that cohesion of the network is determined by certain specific forces: interdependence among agents; a structure of power; a structure of knowledge; a temporary structure (historic development) and spatial structure. Some of these forces are very important to the survival of the network and can be understood as the mechanism that rules the network. We will come to this issue later in this section.

In a different approach, Koleva (2002) distinguishes three types of concepts that are often used as synonyms in network theory: *Alliances*, *Networks* and *Joint ventures*. While joint ventures originate from formal agreements, based on *a priori* specified contractual relationships, networks and alliances emerge from less formalized mechanisms. According to Kogut (2000) they differ due to technological, institutional, or social processes of origin and organization.

In the following, we describe the general idea of transaction costs economics and illustrate its importance in theories concerning networks.

2.4.2. Transaction Costs Economics

According to Mattsson, (1997), the study of organizational networks can be done at different levels of analysis: Micro-level (dyadic relations, that is, on a pair of organizations creating a new partnership), *Meso*-level (network of relationships of one single firm) and Macro-level (network encompasses all the market). Our analysis focuses essentially at the micro-level (since we take into consideration the choice of the partner from a single firm perspective) and at the *meso* level, because it is concerned with networks of firms and their inter-organizational relations.

The study of inter-organizational networks also involves the study of other aspects. One of them is the reason for the existence of organizational networks: *the transaction costs theory*, which will be briefly revised here.

Transaction costs economics studies costs of *negotiation*, *contract achievement,* and *contract attendance*, Williamson (1985). These costs can occur in a market or in the core of a firm (internalised through a set of a processes vertically integrated in a hierarchy).

Markets and hierarchies have been analysed by several authors. Williamson (1975) has distinguished *markets* from *hierarchies* as two different and exclusive alternatives of economic organization, but this vision is not completely clear and accepted, because these forms of economic organization are not seen as totally distinct alternatives and they sometimes overlap. As these forms or organization were thought as having the goal of reducing the costs of a firm, it seemed not reasonable to require that firms always maintain the goal of minimizing the transactions costs. In some situations firms must increase their production or transactions costs when there are reasons to believe that they can obtain some kind of strategic advantages by doing so[13] (Cruz, 1999).

Williamson (1991) has revised his previous work and incorporated the concept of *hybrid form* as an intermediate form of economic organization. *Cooperation agreements* between firms are important intermediate forms of organization that can be identified with these hybrid forms. It has been recognized that there are many intermediate forms of governance that use *trust* and *interdependency* instead of *rationality* (an important attribute in the Transactions Costs Theory) in these intermediate governance structures.

Furthermore, based on Williamson's work, Jarillo (1988, 1993) has studied the economic conditions for the existence of networks and argued that networks can be seen as a form of economic organization that stands between markets and hierarchies. He has

[13] Networking inherent costs can be classified as internal and external networking costs. According to (Ebers, 1997), there is little research concerning the costs that are generally associated with the management of networks.

also suggested that the actions of entrepreneurs can reduce the transaction costs, and establish relations of cooperation (Eiriz, 2004).

2.4.3. Relationships and Networks

Inter-organizational relationships may contribute significantly to the firm performance. Establishing a network relationship requires that the firm should not only desire to form a linkage, but that it should also be attractive to other potential partner agents. Some authors have studied the skills that firms must have to be able to build and manage network relationships that give them a competitive advantage. Ritter (1999) defined *network competence* that is measured by assessing a company's degree of management of qualifications and execution of network activities.

Ritter found four important factors that contribute to the network competence and constitute four antecedents that account for the development and establishment of relationships in a network: *availability of resources, network orientation of human resource management, integration of intra-organizational communication* and *openness of corporate culture*.

The evolution of network relationships changes with the development of firms. Based on a research case study, Lechner and Dowling (2003) identified that every firm has an individual relational mix[14] and that this relational mix changes with the development of firms. The authors studied one *Information & Technology* cluster (the Munich IT cluster) and have identified different types of networks that firms used to realize growth:

- Entrepreneurs, as individuals, have ***social networks*** with individuals from other firms that are important to the business relationships of the firm.

[14] According to Lechner and Dowling (2003) the number of relations within each network and the importance firms give to these networks constitute what they call the *relational mix*.

- Firms have ***reputation networks*** that create future options for relational ties; when an entrepreneurial firm builds reputation networks, it begins to develop its own reputation and moves from being a *reputation-taker* to a *reputation maker*.
- Cooperation with competitors is not uncommon and then ***Co-opetition networks*** emerge, and are important in all stages of a firm's development.
- ***Marketing networks* –** relationships with other firms that enable the central firm to gain better market information for launching new products, to reach new markets, or to gain new clients.
- Knowledge, innovation and technology ***(KIT) networks*** – these networks are relationships that allow firms to have access to technological knowledge and innovation.

2.4.4. Social Networks and Organizational Networks

Social networks are one of the bases of organizational networks. According to Mitchell (1969), "*a social network is a specific set of linkages among a set of persons, with the additional property that the characteristics of these linkages as a whole may be used to interpret the social behaviour of the persons involved.* Social networks are an *entrance ticket* for inter-firm relations, but networks are believed to face quality problems if they are based solely on social relationships (Lechner and Dowling, 2003). Social and reputation networks set the basis for network options in the future but as the firm begins to develop its own reputation, the dependence on reputation networks decreases over time.

Analyzing the persistence of relationships and measuring their survival, Venkataraman and Ven (1998) established a bridge between organizational ecology and network issues. Using temporal data from five start-up firms, these authors investigate certain questions concerning the connection between environmental jolts and changes in the set of relationships of the firms. The authors conclude that:

- Environmental jolts have significant impact on the ending of relationships of a new small firm but not necessary on the creation of new relationships into the same firm. Thus, although liabilities of newness and small size seem to affect the ability of entrepreneurs to hold on to their existing relationships during environmental jolts, these jolts do not seem to affect the entrepreneur's ability to add ties, at least in the very early life of the firm.
- Second, the level of reduction of relationships after environmental jolts seems to be less severe during the early life and greater during the adolescent period. Further entries of new relationships decrease with each succeeding jolt.

Venkataraman and Ven (1998) argue that one implication of their findings is that environmental selection pressures are more severe after an incubation period, called *honeymoon period*. The second implication is the way in which liability of adolescence manifests itself. The *honeymoon* does not provide incubation against exits of relationships. Rather, the incubation manifests itself in the form of the willingness of new costumers and suppliers to join the entrepreneur's transaction set even during a jolt.

2.4.5. Capabilities and Inter-Organizational Relationships

In the industrial sector there are a great number of industrial activities that firms have to execute. However, firms have a tendency to execute activities that are more related to their specialization, according to competences that offer them more competitive advantages (Penrose, 1959).

Within a network, inter-organizational relationships are not isolated because the capabilities that are explored in each relationship are based on other relationships and activities. Besides, inter-organizational relationships are a way to manage the capabilities that firm does not control.

Loasby (1994), studied the bridge between knowledge and capabilities, and stated that most part of the knowledge that a firm must achieve to be successful reside outside the firm. According to the author, firms detain two types of knowledge: the *know-what* (i.e., the formal knowledge that can be codified) and the *know-how,* which is associated with the capacity of developing appropriate actions to achieve a specific goal (Roseira, 2005).

The issue of knowledge transmission in inter-organizational networks leads us to the study of other topics such as collaboration, technological innovation, etc. So, in the following section the issues of cooperation and collaboration are analysed from the perspective of networks. Technological innovation is introduced as a way to promote cooperation between firms, contributing thus to the creation of inter-organizational networks.

2.5. Collaboration Networks and Innovation

In this section, we present an overview of the areas that deal with the concepts of cooperation/collaboration and innovation. Before introducing technological innovation as a key issue in collaborative networks, we discuss the difference between the concepts of collaboration and cooperation (and coordination). After that, we will include the perspective from other authors that classify the types of networks according to their business-level strategies. Finally, after discussing the role of technological innovation in the formation of collaborative networks, we introduce a new networking perspective, coming from the field of virtual organizations and introduce the concept of virtual breeding environments.

2.5.1. Cooperation and collaboration

As introduced in the previous section, collaboration is a fundamental concept in the theory of networks. Collaboration and cooperation can be considered as the complementary action between the elements of a network. *Collaborative strategies* permit firms to specialize in certain activities where they can be more efficient, letting other activities to be accomplished by other members of the network.

Consequently, economies of scale [15] can be attained and costs can be reduced. According to Ebers (1997) collaboration permits firms to enter in inter-organizational networks with the purpose of:

- increasing their revenue by binding competitors as allies and accessing complementary resources and/or capabilities;
- reducing their costs as the result of economies of scale that can be achieved, for example, through joint research, marketing or production (Hakansson and Snehota, 1995). Inter-organizational networking represents *a cost-efficient way of gaining access to crucial know-how that can neither be made available internally nor be easily transferred by licensing.*

Although we have described collaboration as a complementary action between the elements of a network, there are different definitions or views of collaboration and cooperation in literature. Sometimes there is, indeed, not a clear distinction between the concepts of *collaboration* and *cooperation* and they seem to be used indiscriminately.

The work of Wang and Archer (2004) offers a framework that classifies collaboration activities along two dimensions: the level of collaboration and the parties involved in the collaboration. Based in the driving forces and outcomes of collaboration (as *shared risks, information sharing, achieving synergies and seeking competitive advantage), collaboration* is defined as *the "effort made by two or more organizations to achieve results that they cannot achieve working by themselves"* Wang and Archer (2004). On the other hand, in *cooperation*, each party retains its own independence and resources and organizations do not share a common main goal or mission.

[15] Economies of scale characterize a production process in which an increase in the scale of the firm causes a decrease in the long run average cost of each unit.

Winer and Ray (1994) address different levels of collaboration providing a distinction between cooperation, collaboration and coordination:

- In **Cooperation**, each party retains its own authority and resources. Cooperation can take place among parties even when they have not a common goal and a clearly defined and shared mission. Generally, a cooperative relationship involves few risks and little gains. Each party maintains their full independent autonomy and the level of trust[16] between parties is very low.
- In **Coordination** each party retains its individual authority and independence, but specific modifications in the way it operates may begin to occur such as changing work routines, staff task changes, etc. Coordination requires mutual planning and open communication among participants as missions and goals begin to be shared.
- In **Collaboration**, separate parties with commitment to a common mission come together to form a new structure. Collaboration requires comprehensive planning as decision-making, power, authority, and resources are shared. These synergistic efforts often result in innovations that benefit all participants.

Regarding cooperation and collaboration, we will adopt the view above. With respect to cooperation, we assume that there is not a mutual goal (or that this goal is not clear) between the elements of the group, while in collaboration this goal exists.

Coordination will be seen as an intermediary level between cooperation and collaboration and will be associated with of one of these concepts, depending on the type of approach. However, as there is not a consensus in all the literature about this distinction[17], we will avoid a rigid distinction between these concepts, unless this distinction is considered important. Preferentially, *collaboration* will be referred as the generic type of interaction between organizations (that includes cooperation or collaboration).

---

[16] Several authors that analyse the role of trust in organizations. Young and Wilkinson (1989), for instance, analyzed the role of trust and cooperation in marketing channels.
[17] In Industrial Economics, for instance, cooperation is often seen as collaboration or a coordination mechanism.

Cooperation and collaboration are examined again in Chapter 3 when presenting the perspective of Distributed Artificial Intelligence. This will help us to develop our approach that involves Cellular automata and Agent-Based simulations. Meanwhile, we will analyse how collaboration interacts with other important issues as technology, innovation, and networks.

2.5.2. Business-level collaborative strategies

In the literature on business, cooperation and collaboration are often seen as a decision of strategic management. Firms typically join forces following multiple cooperative strategies. According to Hitt *et al.* (2005), a network cooperative strategy covers situations where several firms agree to form multiple partnerships to achieve shared objectives.

Multiple partnerships are particularly important when they are formed by geographically clustered firms (as in Silicon Valley) that maintain effective social relationships and interactions among partners while sharing their resources and capabilities. Besides, one of the advantages of a network cooperative strategy is that firms gain access to their partners' partners (Cline, 2001).

Doing better than competitors (through strategic execution or innovation) and merging and acquiring other companies are the two primary means by which firms develop a cooperative strategy. Following Hitt *et al.* (2005), cooperative strategies represent one of the major alternative firms use to grow. Hitt *et al.* (2005) see strategic alliances as a primary type of cooperative strategy and define strategic alliances as cooperative strategies in which firms combine some of their resources and capabilities to create a competitive advantage.

2.5.3. Technology and collaboration

Technology is a key factor in the economy. With the fast diffusion of technologies, companies unable to face the challenge can rapidly become uncompetitive (Solé and Valls, 1991). If firms are not able to preserve their advantage for long term, technological cooperation may be a possible course of action.

From the technological point of view (De Woot, 1987), cooperation may be used to obtain a technology to overcome the problems of a company which has lost some of its competitive advantages or because the company intends to diversify. From a regional perspective, cooperative forms (as joint ventures) may be seen as strategies for firms to achieve their own aims (e.g. production) in a given geographical territory.

Hakansson (1987) suggests a new perspective that sees innovation (and technological development in general) as a product of exchange among different agents (i.e. a product of a network). The former view treated innovation either as an individual fact (like a Nobel Prize) or as a secret process in an enterprise. In the following we discuss the links between innovation and the formation of inter-organizational networks.

2.5.4. Innovation, Learning, and networks

Innovation is essential to competitiveness and represents *the way in which to anticipate, live with, or react to change* (Ratti, 1991). Innovation capabilities and technological advance have been enhanced by strategic alliances that have proliferated in recent years, particularly in information technology sectors (Gordon, 1991) .

Based on the origin of an innovation, three types of network relations can be distinguished (Kamann and Strijker, 1991): (1) the supplier-dominated relation; (2) the user-dominated relation; (3) the research-dominated network. The question of the origin of the innovation (who gave the impetus) was studied by Laage-Hellman (Kamann and

Strijker, 1991) that supported that innovations are the result of the joint effort of various actors.

As innovation involves learning, Hakansson *et al*. (1999), stress the importance of the business relationships in the diffusion of learning through the network. The more complementary are the organizational competences, the more advantage the firms get from each other.

The authors mention that "*there are reasons to believe that the more connections a relationship has, the greater are the possibilities to learn. […] The argument for this is that if a relationship has a number of connections there are also a number of interfaces where learning could appear: between products, between production facilities and between different backgrounds and competences. A large number of interfaces increase the variation, which is one basic condition for learning*".
The relationship between the connections (network links) and the knowledge stock will be studied in the course of this work, in Chapter 6.

2.5.5. Collaborative Networked Organizations

Collaborative networked organizations (CNO) are a new concept that is emerging in electronic business models. In addition to the rapid evolution of traditional supply chain and outsourcing practices, a growing trend consists of tasks performed by autonomous teams of a small number of people or small and medium enterprises (Camarinha-Matos and Afsarmanesh, 2004). The autonomous teams can be formed by independent firms that are linked by a network, coming together to tackle various projects. They may dissolve once the work is done.

Information and Communication Technologies are frequently seen as an important enabler for CNO's. However, in order to initiate a CNO, the environment where it operates (as well as the organizations and involved individuals) must be previously prepared. Therefore, a long-term base cooperation agreement should be established: it requires the virtual organization breeding environment (VBE).

The breeding environment is a long-term networked structure, having an adequate base environment for the establishment of cooperation agreements, common infrastructures, common ontologies and mutual trust.

Other concepts are also important: V*irtual Enterprise* (VE) is a temporary alliance of enterprises that come together to share skills of competences and resources in order to better respond to business opportunities. The cooperation of the alliance is supported by computer networks. Virtual Organizations (VO's) are similar to VE's except that they are not limited to cooperation agreements in order to obtain profit. A computer network that associates a municipality with other public or private services in order to provide useful information to the citizens with no implicit profit goods is an example of a Virtual Organization.

The establishment of cooperation agreements between firms is not a new phenomenon, but rather belongs to the very nature of the business world. But, as Camarinha-Matos and Afsarmanesh (2004) indicate, "*the use of communication and information technologies to support agile communication as one key characteristic of the virtual concept, brings this approach to a new level of effectiveness. Cooperation on a global scale (intercontinental) is expected to substantially increase, as distance will no longer be a major limiting issue"*.

Organizational networks can be viewed as Virtual Enterprises. Industry clusters or industrial districts are examples of breeding environments in which the cluster is formed by organizations located in the same region, although the physical distance is not the major attribute when cooperation is supported by computer networks.

In addition and according to Osterle *et al*. (1991 p. 2),*"The introduction of transportation technologies and IT, are pushing the physical disintegration of markets* *and enterprises to its global limits and thus enabling the maximal specialization of enterprises*".

In Chapter 6 we will present an application which involves the field of virtual networked organizations.

2.6. Structures and topologies of collaboration networks

This section concerns the structure and topology of networks. After a short foreword, in which we introduce the network structure, we make an overview of certain aspects that are related to collaboration networks, such as *clustering*, *network communities*, *topologies*, etc. After that we identify the main topologies that characterize the most common processes of collaboration.

2.6.1. Key Indicators for Network Structure[18]

Network structure examines how the nodes[19] (actors or agents) are connected (through arcs or links). According to Purchase and Olaru (2003), researchers define some concepts of network structure, such as *centrality*, *clustering*, *density* and *average path length (or connectivity length)*. In his work, Purchase and Olaru examined the effects of centrality within a business network using mathematical models of connectivity and accessibility and centrality applied to transportations networks[20].

---

[18] Jackson and Wollinsky (1996) developed a formal framework that established a paradigm for network structure. This framework was then adopted by many other authors. The main definitions of this framework are presented in Appendix I.

[19] According to literature (e.g. Csermely, 2006) a *node* is a network element with more three links. However, in this work we use the term *node* to refer to any firm in a network, independently of the number of links connected to it.

[20] These indicators aims at evaluating the quantitative properties of a network and do not take into account other qualitative characteristics.

Centrality is one of the issues that are used to determine the position of an agent in a network. Central actors or agents are those who are linked directly or indirectly to a great number of actors than more peripheral agents (Mizruchi and Pots, 1998). Such agents will gain important insights to new advances, scarce resources, and changes within network structure, giving them an advantage over other more peripheral network agents. There are several methods for calculating the centrality. One of the most frequent uses the *Adjacency matrix*[21] (which contains an encapsulation of the most basic connective structure of the network) – including the direct links between nodes.

The *nodality* represents the connectivity of a node and it is computed as the proportion between the number of links connected to the agent and the total number of links. Thus, the higher the value of nodality, the more central is the agent[22].

Clustering is a network property that measures the connectivity inside the network. To quantify this property researchers use the clustering coefficient which is the probability that two neighbours of a node are connected. It is also known as *transitivity.*

Network Density is an indicator of the amount of links in a network. It is measured by the proportion between the number of links and the square of the number of nodes. Sometimes, the *number of nodes* is used computed to measure the network dimension.

The *average path length* or *connectivity length* is the mean number of links that exist in the shortest route between two nodes. Like the average path length, the *diameter* of a network is also a measure of the network length. The *diameter* is the maximum number of links in the shortest route between any two nodes of the network (Csermely, 2006).

Gulati *et al*. (2000) identified these same concepts (such as centrality, density, etc.), and added one other key area of research in which there is a potential for incorporating strategic networks: *structural holes.* In Industry they exist when two trading partners are

[21] The Adjacency Matrix (A) is a square symmetric matrix containing as many rows (and columns) as agents in which each element a_{ij} from A is 1 if there is a link between nodes i and j, and 0 otherwise.
[22] The centrality of the agent provides information about the importance of an agent within a network. However, we must be aware of the simplicity of this indicator, since it does not take into account the importance of the relationships inside the network.

connected only through the focal industry. Structural holes in an industry's customer-supplier network may confer power through control, and ultimately profitability. Based on an analysis of input-output tables that define the network of flows among industries, Burt (1992) has shown that industries that occupy structural holes enjoy greater returns.

In table Table 2-1 we present a summary of the main network indicators, a description of their meanings and a short explanation of how they are measured. Considering the element A (a node or an agent representing a firm), we have:

- *L*: the number of links connected to A;
- *L*: the total number of links in the network
- n: the number of links between all the neighbours of A
- N: the number of links that could possibly exist between all the neighbours of A (a fully connected network between them)
- f: the total number of nodes in the network

Network indicator	Description	How to measure
Centrality	Determines how central in a network an node is	Proportion between the number of links connected to the agent and the total number of links: *l/L*
Clustering (Transitivity)	Measures the connectivity inside the network	Probability that two neighbours of a node are connected: *n/N*
Density	Measures the amount of links in a network	Proportion between links and nodes: *f/L*
Average path length	Measures the network length	Mean number of links in the shortest path between any pair of nodes.
Diameter	Measures the network length	Maximum number of links in the shortest path between any pair of nodes

Table 2-1 - Summary of the main indicators in network theory

2.6.2. Small-worlds and Network Structure

Some authors study certain properties of the network structure that are interesting to analyse. One of them is the *small-world* effect, which was brought to the field of economics by many researchers such as Watts and Strogatz (1997), Csermely (2006), Latora and Marchoiri (2003), etc. Watts and Strogatz (1997) have shown that the connection topology of some biological, technological and social networks is neither completely regular nor completely random but stays somehow in between these two extreme cases. This type of networks is named *Small Worlds* in analogy to the concept observed in social systems by Milgram (1967)[23].

Small world networks are typically highly clustered like regular matrixes but have low path length, like random graphs. These properties are of great importance in economics. This topology is considered more efficient because it is proved to maximize the flow of transmission within the network (while reducing the costs), when propagating information, both at global and local scale (Latora and Marchoiri, 2003).

The work of Leskovec *et al*. (2005) confirm what was said about the low path length of the small-world networks. In their work, graphs densify over time (i.e., the number of nodes increase) and at the same time the average distance between nodes decreases. The diameter of small-world networks decrease as well along time.

Some of these small-world features are explored again later in the case study presented in Chapter 6. In the next section, we will define the topologies that characterize the most common processes of network collaboration.

[23] In their famous small-world experiment, Stanley Milgram gave letters to persons who were asked to pass them to social contacts (friends, colleagues) known on a first-name basis in order to find an unknown distant target (Milgram, 1967). A great number of letters arrived to their destination having gone through a small number of contacts.

2.6.3. Topologies of collaboration networks

Network topology is the study of the arrangement of the elements (links, nodes, etc.) of a network. In economics, the number of topologies observed in cooperation networks is relatively high, from simple to more complex ones. A general overview of the topologies of networks is provided without formal details, the topologies being characterized by three main elements: the firms, the cooperating activities (that are integrated on firms, in our representation) and the flow of resources, as in the ARA model presented above (Hakansson, 1987). Firms and cooperating activities are represented as nodes, and resources as links.

In the topologies given in the figures below, cooperation is associated with buyer-supplier relationships or among buyers or suppliers. In these cases, cooperative relationships are developed for many purposes such as technological innovation, exchange of products and services, etc. The terms cooperation and collaboration remain identical.

We distinguish three main forms that characterize the most common processes of cooperation[24]: *linear*, *star* and *multipolar* networks.

Linear

Within this topology each of the nodes of the network is connected to two other nodes. All the flow that is transmitted between nodes in the network travels from one node to the next node in a linear manner. The network contains different activities, some of which may be managed by different firms.

[24] Several topologies are decribed in literature about collaboration networks. For a more complete overview of topologies based on case studies, one can consult CENESTAP, the Portuguese Center for Applied Textile Studies; this is an important source containing data and case studies about real inter-firm cooperation in Portugal. For the description of topologies in which we based this classification, see Wilhite (2006).

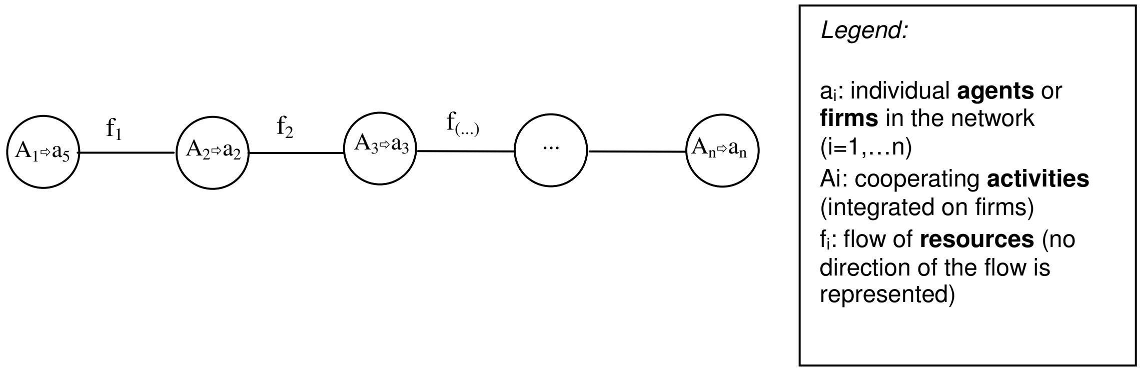


Figure 2-5 – Example of a *linear* network

Note: only some of the flows between nodes are depicted

In this case, the activity A_1 is managed by firm a_1, activity A_2 is managed y a_2 and so on. No direction of the flows is identified in this graph, although the flow of resources can be unidirectional or bidirectional between nodes. Examples of this type of sequential networks correspond to the situations where firms collaborate with partners that are geographically close instead of searching for other networks that are already in action.

Star

The star topology corresponds to the type of network topology in which each of the nodes of the network is connected to a central node. Usually, firms rationalize resources and optimise activities when they form this configuration. The sharing of new technologies from a particular organization and the common benefit of resources are the main advantages of the intervenient of these topologies. Here the flows can also be bidirectional because the sharing of resources is not centralized, but spread into the nodes of the network.

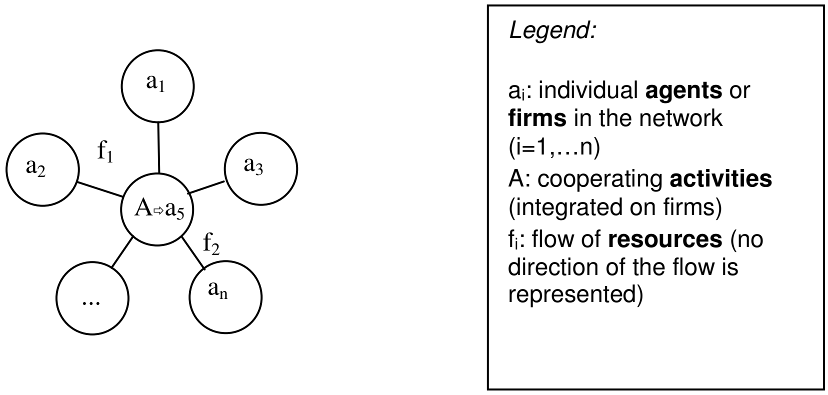


Figure 2-6 – Example of a *star* network

Note: only some of the flows between nodes are depicted

Firms exchance, store or get resources via the central activity A. In many situations new firms are created to aggregate the activities of A: in this case, firm a_5 contains activity A.

One example of networks that follow this topology is the *groups of suppliers*. In this situation, firms are organized around a common client (usually a big dimension client). This form of organizational network represents an interesting opportunity for the cooperating firms to organize more efficiently their supplies. In such a situation the central activity is the goal of the network and constitutes the connection to the final client. The complementarities of the firms' competences are the key for the set of relationships that motivates this type of network, which is based on complementary of vertical relationships[25]. Among the several advantages that firms can take from these *groups of suppliers*, we can emphasize *cost reduction*, *access to new markets* and *risk reduction* in the development of new products.

Faurecia, a world leading automotive supplier specialized in the design and assembly of six majors vehicle modules, can be seen as another example of such topology. Several different suppliers sell directly to Faurecia that integrates the several phases of the production process and contacts buyers (great automobile brands as BMW, Ford, etc) directly. In this case, the network allows for the reduction of costs among suppliers

[25] Firms are said to relate *horizontally* when cooperating with firms in the same market or *vertically* when cooperating with firms that belong to other markets.

because of the common efforts on technological innovation, exchange of products and services, etc.

The topology of star networks is equivalent to another topology named *centralized networks*. Centralized networks emerge usually from entrepreneur's projects that need a centralized structure where resources and activities may be concentrated. Cooperative goals are well defined in this case. R&D alliances use this topology of networks because they locate R&D activities in the central organization. The goal of this modality is well defined and corresponds generally to the reduction of costs and risks. Ho other hand, star networks emerge in situations where is not easy to identify the main goals of the cooperation, and that is the main difference between the two types of networks, although the topologies are similar.

Multipolar networks

The clustering of several organizations characterizes this topology. Typically, the relationships between organizations belonging to the same cluster are strong, but relationships between organizations that belong to different clusters are weak (although the relationships in this latter situation are stronger than with any organization that is outside the network).

Researchers distinguish two different situations in multipolar networks: multipolar polycentric networks and multipolar hierarchical networks. In multipolar polycentric networks, vertical relationships exist (although they do not drive communication), but horizontal relationships between organizations are very important and communication between groups is strong and well organized[26]. Multipolar hierarchical networks contain fragile and fragmented horizontal relationships. Communication is restricted to organizations that are vertically connected.

This kind of networks is typical from the automobile manufacturing sector although they can also be found in other industries, as for instance in textile and clothing industry. Relationships between the cooperating firms from the textile industry and

[26] Horizontally relationships are set up within the same market while vertical relationships with other firms are established between firms belonging to different markets.

firms from the clothing industry can be established with success in multipolar polycentric networks.

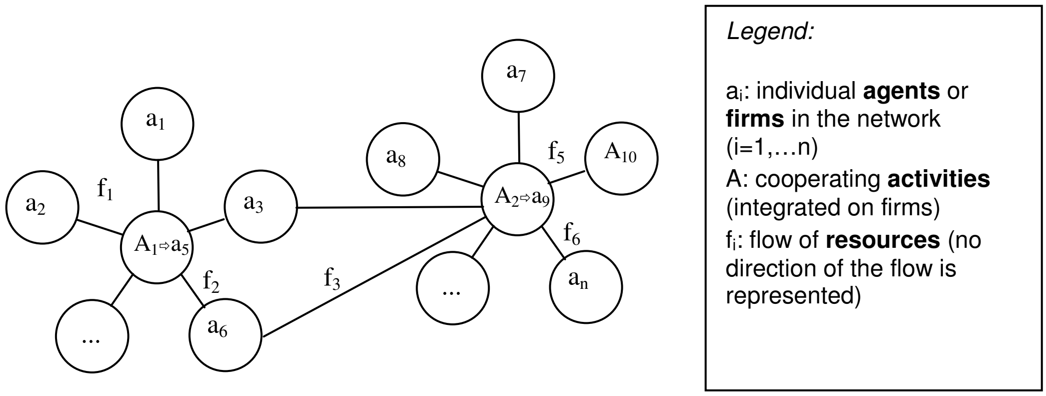


Figure 2-7 – Example of a *multipolar polycentric network*

Note: only some of the flows between nodes are depicted

Figure 2-7 presents such a network containing relationships between organizations belonging to the same cluster, and relationships between organizations that belong to different clusters. In this example two activities dominate the entire network and constitute the reason for collaboration: A_1 (carried out by firm a_5) and A_2 (carried out by firm a_9).

In the next chapter, we introduce some techniques that are used later in the empirical part of our work. We establish the links between individual and aggregate levels of analyses and present the basis of Agent-Based simulation.

3. Interacting individuals and Agent-Based Simulation

This chapter is focused on the links between *individual* and *aggregate* levels of analyses. The connection between these levels is sometimes referred to as *emergence.* According to Axelrod (1997), the large-effects of complex locally interacting individuals endorse the appearance of *emergent* properties at the level of the population.

Some of the concepts that have been discussed earlier in Chapter 2 are now going to be defined as concepts of agent based simulation. Aggregate behaviour may be produced by the interaction of individual agents. Techniques that produce such results are known by several names including Agent-Based modelling, bottom-up modelling and artificial social systems (Axelrod, 1997), that will be studied in the course of this chapter.

A different view is offered by this chapter in which the individual is an abstract entity standing for a real adaptive person or organization (Section 3.1). In Sections 3.2, and 3.3, our main goal is to provide a framework for an empirical application and therefore discuss the simulation, where individuals become *simulated* agents.

3.1. General notions

In this section we will analyse the "*Individual versus Social*" perspective of Schelling (1978), the concept of *adaptive individuals* and *adaptive systems* presented by Belew and Mitchel (1996), and the relation between *learning and evolution* of Young (1998), and Ferber (1999). The investigation of these approaches will facilitate further reading and analysis, where the properties of individuals will be configured in order to provide an aggregate behaviour of the population.

3.1.1. Individual and aggregate Behaviour

Frequently we try to use what is known about individual actions to predict the social aggregate phenomena produced by the group. There are cases in which the aggregate is merely an extrapolation from the individual: if we know, for instance that every driver turns his lights on at sunset, we can guess that the street could appear illuminated almost suddenly. But if most people turn their lights on when some proportion of oncoming cars already have their lights on, we would get a different picture. In this latter case, drivers would be influencing each others' behaviour and responding to each others' behaviour.

This example was given by Schelling (1978), who was interested in studying the macro behaviour that emerged from micro decisions[27]. Some of these micro decisions are influenced by the system of interaction between those individuals and their environment.

[27] As we have presented in the introduction of this chapter, the notion of *emergence* is important since e*mergent behaviour* or *emergent properties* are a consequence of the large-effects of complex locally interacting individuals (Axelrod, 1997).

The notion of behaviour is an important issue. Belew and Mitchell (1996) describe *behaviour*, as something that "closes the loop" between an organism an its environment. The kind of behaviour in Shelling drivers example is *contingent*, that is, the behaviour of some depends on what the others are doing. But this notion of behaviour cannot be isolated from another important concept - the *purpose*. In general, individuals pursue goals, try to minimize his/her effort and maximize his/her comfort. The goals, purposes or objectives relate directly to other individuals and their behaviour.

According to Schelling (1978) the field of Economics has adapted very well to these notions: *"Among the social sciences, the one that conforms most to this kind of analysis is economics. In economics the individuals are people, families, owners of farms and businesses, […]. They are expected to know very little about the whole economy and the way it works. They know the prices of the things they buy and sell, […] and something about the pertinent alternatives to the ways they are currently earning their living, or running their business or spending their money […]. Tens of millions of people are making billions of decisions every week about what to buy and what to sell, or where to work […]"* and the whole system works.

However, the analysis of interactions at the individual level only may lead to misunderstandings. Usually, economists assume that the market is in equilibrium and that the individual preferences are constant. There are examples where this assumption does not hold and the interaction between different levels of analysis is crucial for the real understanding of a given situation. Thus, the analysis of mixed levels (individual cognitive level and social level) may constitute the answer to some of these problems, as we are going to see in Section 3.3.

3.1.2. Adaptive individuals and adaptive systems

Previous paragraphs introduced the issue of individual behaviour and how interactions at this micro level could maintain the whole system working. If we see the interactions in the system as playing in a game, a question arises on whether the game is played always in the same way. In other words, do individuals evolve their strategies? Are they adaptive? Is there any kind of learning behaviour?

The behavioural sciences have focused on the ways in which the behaviour of organisms changes over time at different time scales (Todd, 1996): Moment to moment decision making, long-term alterations of strategies, lifetime developmental adjustments and across life-time changes. These processes are all adaptive, as they involve a change in the behaviour of the individual. However, adaptation requires more than a change in the behaviour, motivated by the environment or by other individuals.

Adaptation *represents the capacity for change and the additional requirement that this change signifies an improvement of fit* (Belew and Mitchell, 1996). Furthermore, changes associated with this improvement of fit (sometimes called *fitness*) must be accumulated and replicated over time (Todd, 1996). The measurement of these improvements is a difficult issue, but some proposals about how to resolve this problem exist in some sciences, such as Economics or Management.

Holland and Miller (1991), has developed the concept of *artificial adaptive agents* (AAA). According to this author an agent is called *adaptive* if it satisfies two criteria: (i) the actions of the agent in its environment can be assigned a value representing, for instance, performance, utility, payoff or fitness, etc., and (ii) the agent behaves so as to increase this value over time.

Adaptation is actually strictly linked to cognitive systems. Herbert Simon (1955) argues that adaptation is a *sine qua non* for any cognitive system and that cognitive science is a fundamental set of common concerns shared by different disciplines concerned with systems that are adaptive. Later in this work, we will simulate some cognitive properties of adaptive agents with the aim of building a Multi-Agent System of organizations. For now, we focus on some of the *cognitive* properties that we can attribute to individual agents in the context of economics.

Adaptive individuals or agents are not irrational. Young (1998) argues that agents adapt but they are not *Hyper-rational*: *"They look around them, they gather information, and they act fairly sensibly on the basis of their information most of the time",* Young (1998, p. 5).

The rationality of economic agents is a very important question in economic science. Simon (1955) developed a behavioural model of *rational choice* where he replaces the global rationality assumed previously in some work in economics by a "*kind of rational behaviour that is compatible with the access to information and the computational capacities that are processed by organisms, including man, in the kinds of environments in which such organisms exist"*.

Simon compares the psychological concept of "aspiration level" of the individuals with the economic concept of the "opportunity cost" in the payoff values. He also discusses the difficulty of humans in being truly rational when all alternatives are evaluated before a choice is made[28]. This approach uses some concepts from Psychology to solve the apparent paradox of the economic theory where there is the attempt to deal with human behaviour in situations in which that behaviour is at least "*intendedly*" rational.

The definition of Holland's artificial adaptive agents discussed earlier was applied to economic theory too. Holland and Miller (1991) argues that "*economic analysis has largely avoided questions about the way which economic agents make choices when confronted by a perpetually novel and evolving world"*. So, an economy can be viewed as an adaptive system, where adaptive agents aggregate. A complex adaptive system, as defined by Holland and Miller (1991) is "*a complex system that contains adaptive agents, networked so that the environment of each adaptive agent includes other agents in the system"*.

[28] Simon, (1955) simplifies what he calls the "classical" concept of rationality where there is some hyper-rationality involving the computation of all possible payoffs involving all alternatives in a decision making process. The author argues that "*there is a complete lack of evidence that, in actual human choice situations of any complexity, these computations can be, or are in fact, performed."*

Coming from the fields of Artificial Intelligence and particularly Machine Learning[29], the theory of complex adaptive systems made it possible to develop well defined and flexible models that exhibit emergent behaviour. The work of Holland (2001 [1975]) introduced a formal framework for the adaptation process. In this framework, there are three associated objects in the centre of the study: the *environment*, *E*, an *adaptive plan*, τ (which determines successive structural modifications in response to the environment), and μ, a *measure of the performance* of the structures in the environment.

Many authors used Holland's and other frameworks to represent their adaptive systems. For example, Carley and Hill (2001), Young (1998) examined adaptation in organizations of intelligent artificial agents. Holland (2001 [1975]) concludes that concurrent learning mechanisms generate the ability to learn strategies which can be either adaptive or maladaptive. Therefore, performance and form of organizations depend on several characteristics, such as *environmental change*, *agent* and *structural learning*, and the emergence of *institutionalized strategies.*

3.1.3. Adaptive Learning and Evolution

The question of *adaptation* that we discussed in previous paragraphs is associated with the issue of *learning*. Indeed, individuals adapt their behaviour in response to others' behaviour and to the environment, in an interactive decision making process where learning plays an important role.

According to Ferber (1999) we can view the problem of the adaptation of a group of agents in two different ways: either as an individual characteristic of the agents – *learning* – or as a collective process bringing reproductive mechanisms into play – *evolution*. It is possible, though, to build an overall view of an adaptation process which is simultaneously individual and collective.

---

[29] Machine learning is the study of computer algorithms that improve automatically with experience (Mitchell, 1997). Genetic Algorithms (GA) lie among the principal learning techniques. GA's were first developed by Holland (2001 [1975]) as a way of studying adaptation, optimization and learning, based on a biological metaphor.

The issues connecting adaptation, learning and evolution have been studied by several researchers including Darwin (1985 [1859]), Lamarck (1914), Baldwin (1896), Hinton and Nowlan (1987) and many others. Their research areas were very different.

Indeed, Psychology and Biology were, *latu sensu,* the original fields of inspiration of those who studied adaptation, but up to date the concepts of adaptive learning and evolution have spread into other fields of knowledge and can be found in almost every subject, from Chemistry and Physics to Economy, Marketing and Computer Science.

Focusing again in the Economic and Social Sciences, Young (1998) systematizes several adaptive mechanisms of learning in the following way:

- **Natural Selection:** Individuals that present higher levels of fitness (well adapted individuals) are at a reproductive advantage compared to individuals that present lower levels of fitness (maladapted individuals).
- **Imitation:** Individuals copy the behaviour of others, especially behaviour that is popular or appears to yield higher payoffs.
- **Reinforcement learning:** Individuals tend to adopt actions that yielded high payoffs in the past, and avoid actions that have yielded low payoffs. This is the standard learning model in behavioural psychology and it has gained the attention of economists. As in imitative models, payoffs describe choice behaviours but it is "*one's own past payoffs that matter, not the payoffs of others*" (Young, 1998).
- **Instance-Based learning:** Individuals use actual information about the environment and other individuals. When a new action has to be adopted by an individual, a set of similar related actions is retrieved from memory and used to decide about the new action to adopt.

It is difficult to say what the best model of learning is. It would be desirable to have a model of learning that incorporates elements of all these models. Table 3-1 presents a summary of the models of learning mainly used by individuals or agents, their advantages and drawbacks.

Models of Learning	Advantages	Drawbacks
Natural selection	Based on natural rules. Seems to work well in several fields of science	Models are based on the idea that rules of behaviour are genetically programmed and poorly adapted rules die out in the long run
Imitation	It is a good approach to acquire effective behaviour	Imitation does not consider one's own choice: only others count
Reinforcement learning	It does not require a large amount of memory to store all data. Individuals just update their knowledge in every iteration.	It takes a very limited view of human rationality: agents are assumed to respond only to their own payoffs; they do not anticipate what others are going to do
Instance based learning	Instead of estimating the target function once for the entire instance space, these methods can estimate it locally and differently for each new action to be classified	Main difficulties include the labelling of new actions to adopt, and the negative impact of irrelevant features on the distance among similar actions that were adopted before

Table 3-1 - Summary of the models of learning mainly used by individuals, their advantages and drawbacks.

One can question whether *natural selection* is indeed a model of *learning*. According to Ferber (1999) and as stated before, if adaptation is seen as a collective process bringing reproductive mechanisms into play, then it is closer to *evolution* than to individual learning. However, in this process there is, in fact, a kind of *adaptive learning* because of the subtle relationship between selection and adaptation that can be summarized as follows: "*Adaptive learning for individuals usually consists of selection among behavioural responses*"[30].

The ecological perspectives presented in section 2.2 are more focused on selection while most of the literature on organizations adopted a different view, which Hannan and Freeman (1977) called the adaptation perspective. The assumption that acquired

[30] As discussed in the previous chapter, learning here plays a role within the organizational adaptation perspective which is closer to the Lamarckian view rather than to the Darwinian one. "*To the extent that learning about the past helps future adaptation, social change is indeed Lamarckian – it transforms rather than selects*" (Hannan and Freeman, 1984). In fact, human actors learn by experience and incorporate learning into their repertoires.

characteristics are not inherited is often taken to imply that the adaptations that an organism learns during its lifetime cannot guide the course of evolution.

However, according to Baldwin (1896) this inference is incorrect[31]. Hinton and Nowlan (1987) hold Baldwins' proposition demonstrating an effect of learning that makes organisms to evolve much faster than their non-learning equivalents during their lifetime. The authors used a neural network to simulate this interaction between learning and evolution.

The evolutionary search of Hinton and Nowlan (1987) was modelled with a version of a genetic algorithm. The authors conclude that some aspects of the environment are *unpredictable*, so it is positively advantageous to leave some decisions to learning rather than specifying them genetically. This argument is different from the Baldwin effect (Baldwin, 1896), which applies to complex co-adaptations to *predictable* aspects of the environment.

The benefits of individual adaptation or "*plasticity*" to evolutionary adaptation seem to be clear: an individual able to adapt to changes in the environment and in response to others' behaviour during its lifetime is surely more fit than one that cannot.

The interaction between the individual and the social levels of learning was studied by Sun (2001). Sun integrated two complementary forms of learning, incorporating two complementary sources of knowledge that an agent possesses: *individual* and *social*. These two forms can interact and the integrated model enables to investigate the influence of socio-cultural processes on individuals.

Littman (1996) stated that many of the problems that allow for an adaptive behaviour may be classified as being part of *evolution* (when they correspond to changes in the

---

[31] The *Baldwin effect* as it has been termed after J.M Baldwin, has its roots in his paper from 1896 entitled "A new factor in evolution". In this paper, Baldwin determines the influence of organic selection and calls it "*a new factor for it gives a method of deriving the determinate gains of phylogeny* [what is observed from the outside] *from the adaptations of ontogeny [*the genetic information*]*". He argues that *"the ontogentic adaptations are really new […] and they are really reproduced in succeeding generations, although not physically inherited"* (Baldwin, 1896). This perspective is similar to the Lamarckian one and different from the one of Darwin.

genome that alter the intrinsic behaviour of the individual) or *learning* (when they correspond to changes in the behaviour as the result of interactions with the environment). Littman developed a set of simulations combining evolution and learning[32]. He has defined two levels: an evolutionary level and a learning level. A genetic algorithm that maintains a population of individuals identified by their genomes defines the evolutionary level.

The learning level takes place during the realization phase of the genomes. For that, a simple neural network was considered. This network codifies the genes of the individual and interacts with the environment. These interactions may originate changes in the individual weights of the neural network and, therefore the network "learns". The changes in the weights die with the individual, but selection occurs at the level of the genetic algorithm.

As an example, the experience of Todd and Miller cited in Todd (1996) on *Exploring Adaptive Agency* is used here. Offspring are born in sub aquatic environment where there are two kinds of substances, identified by their colours that can be consumed: green (food) and red (poison). No information is passed to the new offspring. The goal of their work was to show the connection between learning and evolution processes. They had succeeded in this domain.

In the next section we will describe different types of interactions between individuals, with focus on forms and methods of cooperation.

[32] Complexity rises when the interaction between evolution and learning is considered. Littman used computer simulations to deal with these interactions.

3.2. Interaction and cooperation

3.2.1. Types of interactions

Interaction is a natural process in the real world, resulting from the relationship between an individual and the world where he or she lives[33]. For this interaction to happen, individuals have to be capable of developing communication (not just transmitting information), but more importantly, to provoke some specific behaviour in others (Ferber, 1999). Those interactions may consist, for instance, of attraction, combat, mating, communication, trade, partnership or rivalry (Axelrod, 1997).

Cooperation is the general form of interaction studied most in Multi-Agent Systems, although collaboration and competition are also two important types of interaction between agents.

Under the Distributed Artificial Intelligence (DAI) viewpoint, *the nature of the goals,* the *access to resources,* and *authority* are used to define an initial typology of interaction situations: *Cooperation* is a type of interaction where each party retains its own authority and resources. Besides, there is neither a common goal nor a clearly defined and shared mission. It is a low committed type of interaction and therefore access to resources is not shared and is seen as insufficient.

Cooperation is a mechanism of complex interaction, where agents have to coordinate their actions in the search fort synergic advantages of pooled skills (Ferber, 1999). The level of commitment is low and there is no sharing of goals.

On the other hand, *collaboration* requires comprehensive planning, as decision-making, authority and resources are shared. The level of commitment is high and there is a sharing of goals. These synergistic efforts often result in innovations that benefit all participants.

---

[33] Epstein and Axtell (1996), in their pioneering work called "Growing Artificial Societies", demonstrated how fundamental collective behaviour can emerge from the interaction of individual agents and between agents and the environment.

In the previous chapter, we have seen a similar classification of the types of collaboration from the organizational theory viewpoint (Winer and Ray, 1994). Comparing the organizational theory viewpoint with the present topology of interactions from DAI, this classification presents only some small differences that we describe in the following tables[34]:

(a) DAI viewpoint			
Type of interaction	**Level of commitment to common goals**	**Nature of Goals**	**Nature of Resources**
Cooperation	Low	Compatible	Insufficient
Collaboration	High	Compatible	Sufficient
Competition[35]	None	Incompatible	Insufficient

(b) Organizational Theory viewpoint			
Type of interaction	**Level of commitment to common goals**	**Sharing of Goals**	**Sharing of Resources**
Cooperation	Low	No	No
Collaboration	High	Yes	Yes
Competition	None	No	No

Table 3-2(a) and (b) - Classification of the types of interactions between individuals: the DAI and the Organizational theory viewpoints.

Under the DAI viewpoint, the nature of the goals can be compatible for the elements that interact via cooperation. On the other hand Winer and Ray state that cooperation takes place among parties even when they have not a common main goal and without a clearly defined and shared mission.

[34] We have ignored the *coordination* type of interaction from Winer and Ray (1994), as it stands between *cooperation* and *collaboration*. It will therefore be associated with one of these concepts, depending on the type of approach used.

[35] Within the DAI framework there are several types of competition according to the resources and skills existing in the environment and in the agents. This type of competition is the *collective conflict over resources* that combines other types of competition: collective competition and individual conflict over resources (Ferber 1999).

These two visions are slightly different although this difference does not appear to be significant for a common and definite topology of the types of collaboration or interaction between individuals, as presented in Chapter 2. However, as we have pointed out, this classification will sometimes be ignored.

3.2.2. Forms and methods of cooperation

The reason why cooperation appears in individual interaction has been studied by many scientists. Axelrod (1997) recalls the pessimism of some politicians who argued that before governments existed, the state of nature was dominated by selfish individuals and that cooperation could not develop without a central authority. We know that this is not true today since nations interact without a central authority.

Research in cooperation issues is based upon an investigation of individuals who pursue their own self interest without the aid of a central authority and are forced to cooperate with each other. One of the first simulated experiments dealing with cooperation was the *prisoners' dilemma*, as it was proposed by Axelrod (1984). In this game there are two players. Each has two choices, namely *cooperate* or *defect* and each must take the decision without knowing the others' decision.

No matter what the other does, defection always takes a higher payoff (to the player that defects) than cooperation. The dilemma is that if both defect, both do worse than if both had cooperated, as shown in the following table.

		Player B	
Player A		Cooperate	Defect
	Cooperate	**A=3; B=3**	**A=0; B=5**
	Defect	**A=5; B=0**	**A=1; B=1**

Table 3-3 – Example of players' payoffs in the prisoners' dilemma.

Note: A player takes a decision before knowing the decision of the other player.

It is proved that the payoff is higher when they both cooperate.

This dilemma is an abstract formulation of some general rules which determine when it is best for each individual to go for mutual defection, or mutual cooperation. Axelrod (1984) presents a strategy for playing this game known as TIT FOR TAT. In this policy, players cooperate in the first move. In the following moves, they just do whatever the other player does in the previous move. This strategy was proved to be the more successful in computer tournaments when compared to other strategies, because on average it gives higher payoffs to the players.

This example illustrates how the study of a simple game can contribute to a better understanding of the phenomenon of cooperation. As Axelrod (1984, p. 5) said "*When the goal is to deepen our understanding of some fundamental process [as cooperation], then simplicity of the assumptions is important, and realistic representation of all the details of a particular setting is not*."

Cooperation can be studied from different points of view and under varying conditions. Inspired by the ideas of Ferber (1999) and focusing on the organizational perspective, we have categorized the main objectives for using cooperation - increasing survival capacity, improving performance and resolving conflicts. Besides, we have identified the main methods of cooperation, defining the means that are necessary for the realization of the cooperation. These are shown in Table 3-4.

According to Ferber (1999), the main methods for reaching cooperation are *grouping*, *communication*, *specialization*, *arbitration,* and *negotiation*. The *grouping* method is one of the main methods of cooperation. In this case individuals form a more or less homogenous unit in space or in a virtual network of agents. Agglomeration theories (Location and Urbanization theories) use this method as it is a very important means for organizations to attain shared goals in a physical/virtual delimited space.

Communication means the change of information, while *specialization* is similar to being complementary, as it represents the degree of complementariness that exists in the elements of a group. *Arbitration* and *negotiation* are methods by which the conflict between groups may be solved. In human societies, for instance, the justice system acts as an organ of arbitration.

Main objective for using cooperation is...	Description	Main methods for reaching cooperation
Increasing survival capacity	Cooperation improves collective survival which corresponds to the maintenance of the group	Grouping Communication Specialization
Improving performance	Cooperation contributes for the improvement of performance indicators that translate intrinsic characteristics of the group	Grouping Specialization Communication
Resolving conflicts	Situations of conflict within a group can be solved through cooperation (e.g.: neighbouring organizations living in the same cluster)	Arbitration Negotiation Communication

Table 3-4 – Main objectives and methods for reaching cooperation.

3.2.3. Some related works

Axelrod and Benett (1997) use a *landscape theory of aggregation* to predict how aggregation will lead to alignments among organizations whose leaders are myopic in their assessments. Authors computed and used *distance*, *frustration* and *energy* measures applied to seventeen European nations in the Second World War. Predicted configurations of nations are based on the efforts of actors to minimize their frustration based upon their pairwise tendencies to align with some nations and oppose others.

The previous work of Axelrod and Benett (1997) can be adapted to cover economic applications, namely formation of business alliances. After that, Axelrod (1997) presented a general model for predicting how business firms form alliances to develop and sponsor technical standards.

In their example, the authors started with two simple assumptions: (1) a firm will join a large alliance if it increases the profitability of successfully sponsoring a compatibility standard and (2) the firm will avoid forming alliance with rivals if it is more beneficial to exploit existing compatibility standards. The concept of utility was used instead of profit maximization and showed that the Nash equilibrium is the local minimum of an energy function.

This work is of importance because it combines the integration of a firm in a business alliance with individual utility of that firm, which depends on the presence of rivals and on the size of the coalition. Illustration of the effectiveness of this methodology was made by applying it to the 1998 efforts to create and sponsor UNIX operating systems standards (computers and workstations).

In the next section we introduce simulation and Multi-Agent Systems as these are effective tools to model the interaction among individuals.

3.3. Simulation and Multi-Agent Systems

3.3.1. Simulation as a method

Simulation is a particular type of modelling. A model is a simplification of the world, and a well-recognized way of understanding it. As in a regression equation, simulations also have *inputs* entered by the researcher and provide *outputs*. When a child is playing with her dolls, she is simulating a relationship between a mother and a daughter. Although entertainment may be one of the uses of simulation, there are many other reasons to use this type of modelling. Gilbert and Troitzsch (1999) describe several applications of simulation (some of them overlap) as follows:

- *Better understating of the world* – Simulation is a model of some reality and therefore can be used to better understand this reality;

- *Prediction* – If we develop a model that reproduces the dynamics of some behaviour, we can then simulate the passing of time and thus use the model to predict the future;
- *Provide tools to substitute for human capabilities* - Expert systems have been constructed to simulate the expertise of professionals such as geologists, chemists, and doctors. These systems can subsequently be used by non-experts to predict certain features of the soil or water or to carry-out diagnoses;
- *Training* – Flight or car simulators can be used to train pilots and novice car-drivers; simulation of a national economy may be used to train economists in order to value the impact of changing certain macroeconomic variables;
- *Entertainment* – Games, as flight simulators, are used not only for training, but also for entertainment. Some of these simulations as *SimCity* or *Simlife* are close to social simulations.

In this work we use *computer simulation* as a method of socio-economic research. This method can be associated with the three first reasons defined above. Computer simulation will be used to obtain a better understating of the world, to predict the future, and to provide tools to substitute for human capabilities.

The methodology of simulation assumes that there is some real world phenomenon which the researcher is interested in, called *the target* (Doran and Gilbert, 1994). The idea is to produce a model of this target because it is simpler to study than the target itself. This target is a dynamic entity, changing over time and reacting to its environment.

Methodologically, *computer simulations* can be described in the following way:

i) The researcher identifies the world phenomenon that he/she is interested in – *the target*;

ii) Model development: a model that can be implemented in the form of a computer program is developed based on processes identified in i);

iii) This model is run and its behaviour is recorded;

iv) Evaluation/validation: simulated data produced by the simulation is compared with real data to check whether the model generates outcomes that are compatible to those observed in real world. If the quality of the output would not reach satisfactory level, the process would return to step ii).

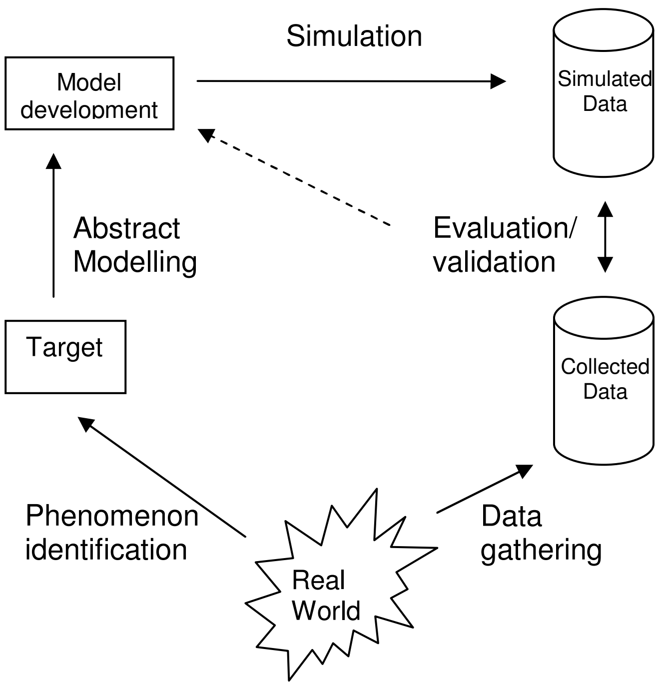


Figure 3-1 - The logic of simulation as a method (adapted from Gilbert and Troitzsch, 1999, p. 16)

Figure 3-1 identifies the main stages of computer simulation-based research. Afterwards we will explore some aspects of this schema, namely the *evaluation/validation* stage, since the comparison of the similarity between real and simulated data is an important issue in this work.

Many simulation approaches[36] use the schema of simulation presented in Figure 3-1, although some of them do not explicitly incorporate evaluation using data comparison. In some studies the outcome generated by the simulated model can be assessed directly by the model builder. Simulated outputs are classified according to a degree of coherence and the model builder accepts the results that maximize the coherence between simulation and real world.

So, it is up to the researcher to choose how the modelling is made. In the following we give an overview of some different approaches to modelling:

- *System dynamics* – This model exploits differential equations. The *target* is described using a system of equations which permits to derive the state of the system at time t+1 from the system state at time t. System dynamics is restricted to the macro level in that it models a part of the reality (the *target* system) as a whole. One application of system dynamics is WORLD2 (Forrester, 1971) implemented in a software called Dynamo[37]. This application is a world model where pollution, population, natural resources and the capital stock sector can all be modelled.
    - *Queuing models* – These models are discrete event models. Events are scheduled using an *agenda* and events can be predetermined to start at a given time. There are different kinds of objects in queuing models, namely *servers*, *customers, queues and the agenda*. One of the applications of queuing model was implemented in *SimLab*[38] and aims at simulating an airport. It includes costumers (aircraft, passengers), servers (gates, runways), etc.
    - *Micro-simulation* - Modelling is produced at the individual level. A set of rules (transition probabilities) are then applied to these units leading to

[36] Gilbert and Troitzsch (1999) present a detailed analysis of the main simulation techniques used in social sciences.
[37] The simulation language Dynamo (DYNAmic MOdels), was developed in the late 1950s by a group working with Jay Forrester at MIT (http://web.mit.edu/sdg/www/).
[38] SimLab is the multidisciplinary research unit at Helsinki University of Technology, Computer Science and Engineering, Information Networks Study Programme. The core of SimLab is a virtual collaborative learning environment for business processes.

simulated changes in state and behaviour. These rules may be deterministic or stochastic. Stochastic micro-models are opposed to the (usually deterministic) macro model of the system dynamics approach. This type of simulation is used, for example, to model the social security system of some country, by modelling the evolution of the attributes of individuals such as births, deaths, age, retirement, etc. (Davies and Favreault, 2003).

- *Cellular automata*[39] - Cellular automata are cells located in a regular grid where the behaviour of an individual cell is determined by a set of rules which specify how that state depends on the previous state of that cell and the states of its neighbours. The Conway's Game of Life (Gardner, 1970) was implemented with this simulation technique. The rules are simple: a cell can only survive if there are either two or three other living cells in its immediate neighbourhood (the eight cells surrounding it). If these conditions are not satisfied it will die either from the effect of overcrowding (if it has too many living neighbours) or from loneliness (if it has too few living neighbours).
- *Multi-Agent Systems* – Although there is not a general consensus about what an *agent* is, the term is usually used to describe *self-contained programs which can control their own actions based on their perceptions of their operating environment* (Huhns and Singh, 1998). There are several applications of Multi-Agent Systems in social and economic sciences. For instance, Carley and Hill (2001) created *Construct-O*, a Multi-Agent model of social and individual change resulting from the diffusion of information among adaptive and communicative individuals. In this model, each agent has a certain position in the social network, together with a mental model or knowledge. When individuals interact, they communicate a piece of their knowledge so information diffuses.

[39] Cellular automata are - by definition - dynamical systems which are discrete in space and time - and are characterised by "local" interactions. Cellular automata are an extension of a concept often used in Artificial Intelligence: *automata*. The, *Automata theory* is the study of abstract machines and problems they are able to solve. A single automaton is a mathematical model for a finite state machine (FSM), which is a machine that, given an input of symbols, jumps through a series of states according to a transition function (Hopcroft and Ullman, 1979).

The following table compares these simulation techniques according to some of their features, namely *communication*, *complexity*, *advantages* and *drawbacks*.

Simulation techniques	Communication between agents	Complexity of agents	Some Advantages	Some Drawbacks
System Dynamics	No	Low	Clear formulation of the problem; analytical solutions may exist	Difficult to change attributes during the simulation
Queuing models	No	Low	Focuses in the notion of the **event**	Customers do not change the behaviour during the simulation
Micro-simulation	No	High	Stochastic simulation with possibility of changing of individual attributes	Few levels of analysis. Individuals have no autonomy
Cellular automata	Yes	Low	Introduces communication between agents	Agents are very limited in their behaviour and intelligence
Multi-Agent Systems	Yes	High	Agents are given the expression of human or organizational activity	---

Table 3-5 – Main social simulation techniques and some of their features

We did not find any general disadvantage in Multi-Agent Systems except that in most part of the cases no analytical solutions exist. We could point out that Multi-Agent Systems may be quite complex and this may complicate the understanding of the output. Besides, Multi-Agent programming can answer some questions that arise in the fields of economics and game theory. Some scientists, nowadays known as the founders of computer science, such as Von Neumann and Alan Turing, were interested in the formal study of games. It may be that this interest (that ultimately led them to write their classical papers) is seen as the foundation of Artificial Intelligence (Wooldridge, 2002).

In addition, simulation using Multi-Agent Systems has turned out to be a very useful technique and is rapidly becoming important in every scientific field because of their simplicity and efficiency when run on existing computers. Agent software has been much influenced by the work of Artificial Intelligence, (AI) especially the subfield called Distributed Artificial Intelligence (DAI) (Bond and Grasser, 1988). DAI is very important to social simulation, because it pays attention to building networks of intelligent agents and investigating their properties.

Multi-Agent Systems have many designations as *Agent-Based modelling*, *Bottom-up modelling* and *Artifical Social Systems* (Axelrod, 1997); all aim at analysing the aggregate behaviour (and to understand complex properties) arising from the interaction of the simulation among the individual agents.

Other authors use different terminology. Holland (2001 [1975]) introduced the concept of *Artificial Adaptive Agents* (AAA). According to Holland (1991) AAA *models show some characteristics that are not available in traditional modelling techniques* […]. *Mathematical models lose flexibility but gain a consistent structure and general solution techniques.* […] *The resulting [AAA] models are dynamic and are "executable" in the sense that the unfolding behaviour of the model can be observed step-by-step. This makes it possible to check the plausibility of the behaviour implied by the assumptions of the model.*

Axelrod (1997) argues that Multi-Agent Systems are an intuitive method that contrasts with the standard methods of induction and deduction: *"Agent-based modelling is a third way of making science. Like deduction it starts with a set of explicit assumptions. But unlike deduction it does not prove theorems. Instead, an Agent-Based model generates simulated data that can be analyzed inductively. Unlike typical intuition, however, the simulated data come from a rigorously specified set of rules rather than a direct measurement of the world. Whereas the purpose of induction is to find patterns in data and that of deduction is to find consequences of assumptions, the purpose of agent-based modelling is to aid intuition."* (Axelrod, 1997, pp. 3-4). This intuition referred to by Axelrod (1997) in often linked to *emergence.*

As other simulation techniques, Agent-Based modelling is a way to do experiments. But *"agent based modelling is particularly useful when the population is heterogeneous, when interactions among agents are complex and nonlinear, and when the space is crucial"* (Epstein, 2003). The large-effects of complex locally interacting agents are called *emergent properties*. Epstein and Axtell (1996), in their pioneering work, demonstrated how fundamental collective behaviour can emerge from the interaction of individual agents.

In the view of Holland (2001 [1975]) and Holland and Miller (1991), emergence is one of the effects of complex systems. The author argues that Multi-Agent organizations are complex systems, and many economic systems can be classified as complex adaptive systems. Such a system is complex in a special sense, because:

(i) it consists of a network of interacting agents (processes, elements);

(ii) it exhibits a dynamic aggregate behaviour that *emerges* from the individual activities of the agents;

(iii) its aggregate behaviour can be described without a detailed knowledge of the individual agents.

In the next sections we examine Multi-Agent Systems (MAS) in more detail. We analyse the characteristics of agents and MAS, including their cognitive aspects, and discuss some applications, as well as the problem of *validation*.

3.3.2. Multi-Agent Systems

Let us clarify and systematize the concepts that have been introduced in the previous sections:

- **Interaction -** *An interaction occurs when two or more agents are brought into a dynamic relationship through a set of reciprocal actions;*

- **Behaviour -** *characterizes all the properties that the agent manifests itself in its environment;*
- ***Environment -*** *is the place where agents "live": it can be a centralized environment, or a distributed environment;*
- ***Communication:*** *is the basis for interactions and social organization; it is expressed as a form of interaction in which the dynamic relationship between agents is expressed through the intermediary of mediators or signals, which once interpreted, will affect the other agents.*

In the next section the characteristics of agents, as well as Multi-Agent Systems and examples that can be modelled by the means of MAS are discussed in more detail.

3.3.2.1. Characteristics and Modelling of Agents

In the Multi-Agent Systems' literature, *an agent* is usually seen as an entity that lives in an environment and that is capable of interacting with other agents. *Agents* are frequently characterized as having the following characteristics (Ferber, 1999):

- ***Action*** *and* ***interaction*** – Agents interact with other individuals and with the environment. Action modifies the agents' environment, and thus future decision making;
- ***Communication*** *with other agents* – the main way in which agents interact;
- ***Individual goals*** and ***Autonomy* -** agents are not directed by commands coming from a user or another agent, but by a set of "tendencies" which can take the form of *individual goals* or *satisfaction* or *survival functions* that agents try to maximize;
- ***(Limited) Perception –*** agents have only a limited or partial representation of the environment where they live in. In other words, they have no overall perception of everything that is happening around them. It is often assumed that

they have "bounded rationality" in the sense that they use limited computational resources to derive the consequences of what has been perceived[40].

Wooldridge (2002) and Gilbert and Troitzsch (1999) have defined a different, somewhat complementary classification of agents' characteristics which include reactivity, proactiveness, social capability and autonomy (Gilbert and and Troitzsch, 1999). In the following we will examine the issue of reactivity. Regarding the other issues the reader can consult the references given.

According to Ferber (1999) there are two schools of thought based in two concepts: *cognitive* and *reactive* agents. In the school that includes cognitive agents, a Multi-Agent System contains a small number of agents in which each agent's knowledge is represented in the form of a separate knowledge base.

The Reactive school asserts that for the system to demonstrate overall intelligent behaviour, it is not necessary for individuals to be individually intelligent. Societies of ants (Corbara *et al*. 1993) offer examples of organizations of this type. All the ants maintain the colony without having cognitive capacities.

In the following we will focus our attention on cognitive agents that are used later in our simulations.

3.3.2.2. Representing mental attitudes of Agents: BDI architecture

According to Tecuci (1998), an intelligent agent is a knowledge-based system that perceives its environment, rationalizes, draws inferences, solves problems, determines actions, and acts upon that environment to realize a set of goals or tasks.

The behaviour of the agent is based on a correspondence between certain aspects of the environment (domain) and its internal model. This internal model includes a set of

[40] As presented above in Chapter 2, Herbert Simon developed a behavioural model of rational choice (Simon, 1955) in which he considers a "*kind of rational behaviour that is compatible with the access to information and the computational capacities that are processed by organisms*".

desires, *beliefs* and *intentions,* a *reasoning component,* containing a *knowledge base*, a *learning engine* and an *inference engine*.

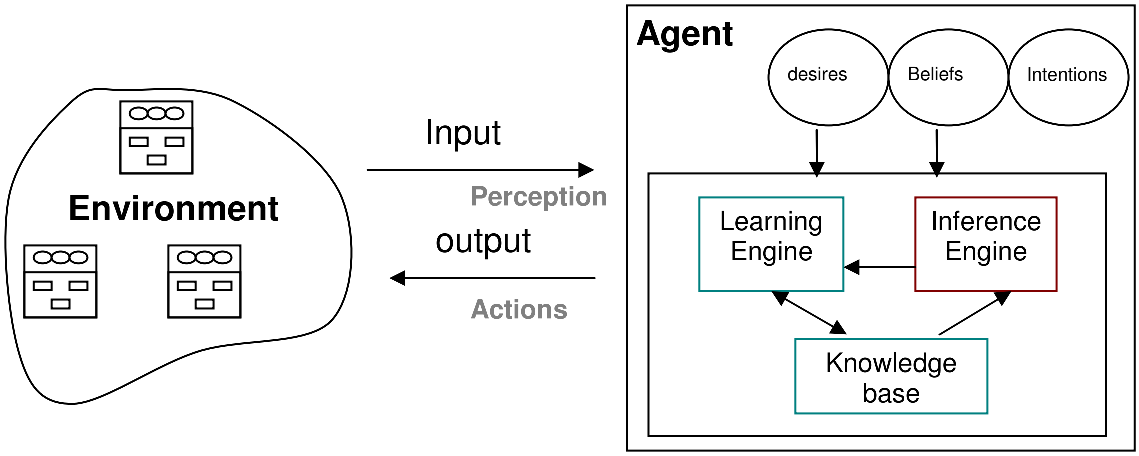


Figure 3-2 – Architecture of a cognitive agent with emphasis in the learning aspects, based in Tecuci (1998)

Figure 3-2 represents the *model* of a learning agent, interacting with the environment (that contains other agents). In this model the aspects of agent cognition are very simplified. In fact, mental configuration involves many types of concepts, including *beliefs*, *intentions* and *desires*; but there are others as *percepts*, *emotions* etc.

This configuration of an agent follows the so-called BDI architecture (Beliefs-Desires-Intentions), developed by Rao and Georgeff (1991). *Beliefs* correspond to a component of an agent that embodies the information (states of the system[41]) and which is updated appropriately after each action of the agent.

Desires are goals or objectives (i.e. what the agent desires as the final state of the system) corresponding to the motivation of the agent[42]. *Intentions* represent the currently chosen course of actions intended to be taken to achieve the desires. In other words, intentions can be seen as a sequence of actions that the agent takes in order to

[41] If we take the environment as a system, the term *state* describes any overall configuration of a system. This concept will be used later on this work together with Multi-Agent Systems.
[42] Rao and Georgeff (1991) distinguish *desires* from *goals. Desires* may be associated with some instant of time and may be incompatible, while goals are usually permanent and unique in a certain period of time.

achieve its desires. *Intentions* play an important role in practical reasoning and can not be reduced to beliefs and desires. *Desires* capture the deliberative component of the agent (Rao and Georgeff, 1991). So, the BDI architecture views an agent as someone that is rational and has some certain mental attitudes of Belief, Desire and Intention representing, respectively, the informational, motivational and deliberative conditions of the agent.

This architecture has been criticised by many researchers, including Rao and Georgeff (1995). Some say that it is very reductive: sociologists and scientists from Distributed Artificial Intelligence argue that there are other cognitive aspects that should be introduced in the model. On the other hand, classical decision theorists question the necessity of having all these three attitudes. Other authors argue that it is hard to find a mechanism that permits an efficient implementation of the mental attitudes: *beliefs*, *desires* and *intentions.*

Anyhow, the BDI-architecture has been used in many applications and it seems to be adequate in dealing with many situations. Furthermore, the concepts of *beliefs*, *desires* and *intentions* are easy to understand and the BDI–architecture has the advantage that it is intuitive and relatively simple to identify the process of decision-making and how to perform it (Corchado *et al*. 2004)

In Figure 3-2, the agent is endowed with to the ability to learn as it contains a learning engine that is capable of creating and updating the data structures in the knowledge base. Learning agents will be therefore considered as agents that are able by themselves to acquire and maintain the knowledge (Tecuci, 1998).

Adaptation of an individual depends on the way it gets and assimilates the information from the environment, as we saw in Section 3.1. Information is subject to error and, also, it is the result of the agent's perception of the environment. Agents having this capacity of being able to call into question a fact develop the basis of our cognitive adaptation and therefore the accommodating of our cognitive system to a world of perpetual evolution (Ferber, 1999).

Leaning is an important aspect in the agent architecture. From a philosophical point of view, according to Ferber (1999), all learning relates to beliefs. "*It is always a matter of information obtained from something else or someone else"* Ferber (1999, p. 245). The model that is shown in Figure 3-2 includes a learning engine, a knowledge base and an inference engine.

In this model the way learning works is shaped by the *beliefs* and *desires.* In fact this kind of influence happens in most of the learning situations. Learning in human beings depends on beliefs and desires. For instance, if the user knows that some barometer indicates the opposite air pressure (and not the real one), the user will conclude that it will be a nice day, when low pressure is shown. So, the user took his representation of information (a kind of a belief) to interpret the barometer.

BDI-architectures have been used in many applications. For instance Corchado *et al.* (2004), used it in an agent-based application of a wireless tourist guide combining the BDI approach with the learning capabilities of case-based reasoning.

3.3.2.3. Characteristics and Modelling of Multi-Agent Systems

The term Multi-Agent System (MAS) is applied to a system exhibiting an architecture containing the following elements (Ferber, 1999; Wooldridge, 2002):

(i) An environment E;

(ii) A set of objects, O situated in E;

(iii) A set of specific objects, called agents, A such that $A \subseteq O$;

(iv) An assembly of relations, R that links objects to each other;

(v) An assembly of actions, Op, making it possible for the agents of A to *perceive*, *produce*, *consume*, *transform* and *manipulate* objects from O;

(vi) Operators (or functions) with the aim of representing the application of the operations (*action* and *behaviour*).

Here we distinguished architecture from the behaviour of agents: (i) the *architecture* deals with the question of agent integration in the MAS; (ii) the *behaviour* deals with what an agent does and how he interacts with other agents and/or with the environment.

Multi-Agents are implemented using computers, so a set of types of languages have been defined to design a MAS. According to Ferber (1999) there are five types of languages in the MAS design[43]:

- **L1** - used for the implementation of a MAS – e.g. JAVA, C++;
- **L2** - communication languages as ACL and *KQML*[44];
- **L3** - languages for describing actions and behaviours and the laws of the environment;
- **L4** - Languages for representing knowledge which are important when agents are very complex; e.g. *blackboard-based* languages[45];
- **L5** - Formalization and specification languages at the most abstract level dealing with cognitive aspects as beliefs, intentions, etc.

This classification is organized from the lowest level (implementation) to the highest one (abstraction). This representation has been defined in the field of Artificial Intelligence (Ferber, 1999) where states and actions are represented with the help of

[43] Although these types of languages do not need to be alike, they must necessarily be compatible with the same implementation type.

[44] **KQML** is the acronym for *Knowledge Query and Communications Language*, a standard high level communication language (from DARPA Research Project – U.S.) based on speech acts to allow cognitive agents to cooperate. **ACL** (Agent communication language) is another standard funded by FIPA (Foundation for Intelligent Physical Agents) to face some of the existent drawbacks of KQML as the ambiguity and imprecision due to the natural language descriptions.

[45] Blackboard-based languages are the most widely used symbolic cognitive architecture for Multi-Agent Systems. They are based on three elements: *Knowledge Sources* (modules that are used to share data), a *Shared Base,* and a *Control device* that manages conflicts between the *Knowledge Sources.* Engelmore and Morgan (1988) offer a deeper explanation of these Blackboard-based languages.

interconnected components. An overview of this framework is presented in Appendix II.

In a different perspective, based on empirical validation, Fagiolo *et al.* (2005) define that the minimal structure of a MAS shall contain the following elements:

Time (t = 0, 1, 2, …): e.g. quarters, years, etc.;

Set of Agents (A_1, A_2, …, A_n) ; e.g. consumers, firms, etc.;

Micro Decision Rules, $R_{i,t}$;e.g. production rules, microeconomic rules, etc.;

A set of *variables* X that measure the behaviour of Agents (X_1, X_2, …, X_k); e.g. agent's Profit, Marginal cost, etc.;

Vectors of *Micro-States*, X_i,t (individual variable with time); e.g. profit of agent (X_k) at time t ($X_{k,1}$, $X_{k,2}$, …, $X_{k,N}$);

Aggregate Variables at time t, $X_{kt} = f (X_{k,1,t} , …, X_{k,N,t})$, (aggregate values of variable X_k at time t); e.g. profit of all agents at time t; birth rate of a population of firms, etc.;

Vectors of Micro-Parameters, θi ; e.g. number of initial firms in the beginning of the simulation;

Vector of Macro-Parameters, Θ; e.g. number of total firms in the simulation;

Interaction Structures, Gt ; e.g. networks.

In studying MAS, Fagiolo *et al.* (2005) found out that the micro and macro variables are governed by complicated processes which can hardly be analyzed analytically because of the non-linearities and randomness that emerge in individual behaviours and interaction networks. According to the authors, that is why many Agent-Based models are generated by the means of computer simulations

The question of the empirical validation of MAS deals with some of the elements referred to above. We will come to this issue again later in Section 3.3.3 and in Chapter 4.

3.3.2.4. Integrating Cognitive and Social levels in Modelling

Several researchers studied the role of cognitive models in social simulation. Gilbert (2005) suggested that *"it is possible to distinguish at least a biological, a cognitive and a social level* [of analysis]*, in which the characteristics of phenomena at one level are emergent from the behaviour of phenomena at levels below"*.

It is important to note that different levels of analysis need not be present in the study. Some social simulation models were elaborated and these did not include the cognitive level. However, there can be important interactions between levels which should not be ignored. Consider, for instance, a cognitive level and a social level that can be used to increase the coordination of the agent's actions. That is what happens, for instance, in case of flocking birds, when this aggregate behaviour can be the result of the interaction of individual and social factors.

Reynolds (1987) and Heppner and Grenander (1990) presented simulations of various movements in a bird flock or fish shoal. Both studies had the insight that local processes might underlie the unpredictable group dynamics of bird social behaviour. One motive for developing such simulations was to extend this model to human social behaviour. However, humans adjust not only physical movement (as the individuals in flocking birds do), but cognitive or experimental variables as well (Kennedy and Eberhart, 1995). Humans tend to adjust their beliefs and attitudes to conform to their social peers.

There are optimization methods based on the simulation of these social models of bird flocking and fish shoaling, such as the so-called Particle Swarm Optimization (PSO) developed by Kennedy and Eberhart (1995), and improved by Hu *et al*. (2004). PSO is a well known method used for optimization of continuous linear functions incorporating some aspects of genetic algorithms and evolutionary programming. PSO is initialized with a group of random particles (solutions) and then searches for optima by updating each generation. Each particle, in every generation is updated by following the two "best" values (the values with best fitness).

Another model that explores the interaction of individuals in a social system is the *Ant Colony System*. In this approach, the "ants" are agents with very simple basic capabilities which mimic the behaviour of real ants. In real world the moving ants lay specific pheromones (a medium used to communicate among animals) on the ground, thus marking the path by a trail of this substance. This can be simulated in a simulated world. An isolated ant moves essentially at random, but an ant that encounters a previously laid trail can detect it and decide with high probability to follow it, thus reinforcing the trail with its own pheromone (Dorigo *et al*. 1996). The collective behaviour emerging in this process has the following characteristic: the more ants follow a trail, the more attractive that trail becomes for others. The process is characterized by a loop where the probability with which an ant (agent) chooses a path increases with the number of ants that previously chose that path.

The use of individual and social cognitive models has been helpful in developing better models of aggregate processes in social and economic theories. Concerning cognition, Sun (2001) identifies two sources of knowledge that an agent possesses based on their origins: individual and social[46]. Sun developed his *Clarion Model*, consisting of two levels: the top level - conceptual - and the bottom level – behaviour oriented. These two levels interact by: (i) cooperating in actions, through a combination of the action recommendations from the two levels; (ii) cooperating in learning through a bottom-up and a top-down process. The author used a learning plan where at the top level knowledge is captured in a simple rule form, while at the bottom level *reinforcement learning* appears in different stages with the implementation of Q-learning algorithm.

According to Sun (2001), cooperation in learning through a bottom-up and a top-down process, is just the combination of *assimilating social structures* and relates *learning by being told* (Michaslki, 1983) to top-level learning, and *autonomous learning,* to the bottom level of the reinforcement learning.

3.3.2.5. Networked Agents

[46] According to the author, these are extreme cases because there are many situations in which knowledge involves both sources.

Many Multi-Agent applications use networks as a support for their activities. Agents can be viewed as the nodes of the network with the arcs or links indicating some kind of relationship between the agents. In the previous chapter, networks have been studied within an organizational perspective: networks are systems of relationships based on a division of work in the network.

The study of networks has improved considerably with the help of Agent-Based simulation. According to Purchase and Olaru (2003), simulation has been a favourite approach in the study of networks due to its capacity of exploring *emergent behaviour* as a result of the interactions within the network.

Simulation is a useful tool for the analysis of the dynamical behaviour of networks (mainly used for creation and deletion of links). There are many examples of network applications developed with the use of agent-based simulation: Pujol *et al*. (2005), Wilkinson *et al*. (2001), Gans *et al*. (2004), Ormerod *et al*. (2002), Newman (2000), Axelrod (1997), Epstein and Axtell, (1996), etc.

3.3.2.6. Agent-Based Computational Economics (ACE)

Multi-Agent Systems are often used to model economies and organizations. According to Tesfatsion (2006), there are several reasons why economies should be modelled by MAS: economies are complicated systems encompassing micro behaviours, interaction patterns, and global regularities. The MAS allows a "*two-way feedback between microstructure and macro regularities mediated by agent interactions*".

Whether partial or general in scope, studies of economic systems should consider how to handle difficult real-world aspects such as asymmetric information, imperfect competition, strategic interaction, collective learning and possibility of multiple equilibria.

One of the most important aspects in modelling in economics is the way how the system is conceptualized and then implemented by the modeller. According to Tesfatsion (2006), the modeller constructs a virtual economic world populated by various agent types (economic, social, biological, physical) and sets the initial world conditions. Next, he/she *steps back* to observe how the virtual world develops over time. The events in this virtual world are driven by agent interactions.

Tesfatsion also identifies the key characteristics of an organizational agent, what she calls the "firm" agent. It should contain the following characteristics:

- Profit-seeking agent with strategic reasoning and learning capability;
- Profit gains by producing and selling products;
- Can adjust production and price levels in every trading period;
- Can also invest profits to expand its production capacity.

Multi-Agent Systems are a way of "growing economies from the bottom up". The so-called new field of *Agent-Based Computational Economics* (ACE) is inspired, includes or may be complementary to other fields such as: Experimental Economics, Complex Adaptive Systems, Computational Modelling of Social Dynamics, Network Economics, Agent Modelling in Economics and Finance, etc.

There are many applications of Multi-agents in the Economics and Social Sciences. Typical applications of MAS are the modelling of *populations of organizations in the managerial sciences*, *social simulation,* including *socioeconomic modelling*, more specific problems of strategic management and economics of networks, etc.

Examples of the use of MAS that model cooperative behaviour include the work of Ant Colonies, as that of Dorigo *et al*. (1996), and some specialized topics such as the work of Andras *et al*. (2006) who studied the mixed effects of uncertainty and cooperation. Uncertainty is an important factor that influences social evolution in natural and artificial environments; environmental uncertainty permits to capture the variance of the resources in the environment. Authors conclude that cooperation can protect against the unfavourable effects of uncertainty for the individual.

3.3.2.7. Virtual breeding environments and MAS

As introduced in section 2.5.5, virtual breeding environments (VBE's) are long-term networked structures, presenting the adequate base environment for the establishment of cooperation agreements in computer networks. VBE's are real examples of MAS' applications. As pointed out by Camarinha-Matos and Afsarmanesh (2004), there is actually an analogy between Multi-Agent Systems and virtual enterprises. Among other aspects, this similarity includes the following aspects:

- Each organization can be considered as an autonomous unit, as an agent;
- The creation of a VE (Virtual Enterprises = temporary alliance of enterprises) is analogous to the creation of networks or coalitions under the perspective of a MAS.
- The processes of VE creation are based on negotiation rules and scenarios similar to the protocols existing in the MAS domain.

Furthermore, the use of MAS in virtual breeding environments has many advantages including the possibility of predicting the behaviour of the VE and linking directly the simulation with real emergent processes.

In Chapter 6 we describe a Multi-Agent System that has been inspired by real examples as *e-marketplaces* that we can consider as a particular case of a Virtual Enterprise.

3.3.3. Questions of validation in MAS

The comparison of real and simulated data can be regarded as an evaluation stage. It is an important step that helps to validate the simulation. Computer programs can produce erroneous results and validation is needed to determine whether the output is of any value.

Empirical validation of Agent-Based (AB) models is the main concern for researchers that use simulation models in social sciences. In the last two decades researchers have improved the quality of simulations, namely in economical modelling by extending their modelling framework to incorporate certain aspects that have been missing until then: agent heterogeneity, bounded rationality of the agents (not hyper rationality), learning and technological change (Windrum *et al*. 2007).

The need for empirical validation is, nowadays, essential in model's evaluation. In economics (and in other sciences), the basic methodology to built AB simulation models consists in isolating some features of a particular phenomenon in order to understand it and to predict its future states under new conditions.

Fagiolo *et al.* (2005) have identified different methods of empirically validating Agent-Based models:

- *Replication of Stylized-Facts:* in this case, empirical validation is done at the aggregate - macroeconomic – level, and parameters and initial conditions are not restricted a priori; validation requires joint reproduction of a set of "stylized facts". The authors refer to this method as *indirect calibration;*
- *Empirical Calibration of ABMs*: this method deals with the space of initial conditions and micro/macro parameters and uses empirical knowledge to calibrate initial conditions and micro/macro parameters;
- *History-Friendly Industry Models*: models built upon detailed empirical historical knowledge of phenomenon under study and employed to replicate its precise (qualitative) history; the goal is to empirically validate the model by comparing "simulated trace histories" with "actual history" of an industry.

The historic-friendly approach is strongly quantitative and mainly focuses on micro-economic transients of industrial paths of development. It implies detailed historical paths corresponding to the data being analysed. The second approach (*Empirical Calibration of ABMs*) requires a good empirical knowledge of the initial conditions and micro/macro parameters.

The Replication of Stylized-Facts requires a set of summarized evidences that can be used to calibrate the model. Sometimes these evidences are formalized in a specific objective function that can be used to optimize the model. That is the case of Rogers and Tessin (2004), which used an Agent-Based model of a financial market and calibrated it using a multi-objective genetic algorithm.

The analysis presented so far permits us to define the research questions that we have set out to investigate. In the next chapter, we define the main questions and the strategy adopted for our work.

4. Research Questions and Methodology

In this chapter we define the main research questions and the strategy adopted for our work. First, in Section 4.1, we identify the problem and the main research questions. So far, we have presented several different perspectives to deal with the problem of the adaptation and survival of the firms. As these different approaches are complementary, we will benefit from the cross-fertilization that exists among them and exploit them to answer to the main questions proposed in this chapter.

In Section 4.2, we introduce the general research strategy. These issues are discussed in the following chapters and then tailored for each of the studied applications. The approach adopted is then evaluated. The aspects of the original contributions of our work are presented in Section 4.3. An overview of the state of the art in the resolution of similar problems and related works are given in Section 4.4., and finally in Section 4.5 we present some empirical evidences of collaboration that support the specific hypothesis that will be tested in each of the studies presented later in Chapter 5 and Chapter 6.

4.1. Problem Definition and Main research questions

Important aspects to retain from literature

Organizations interact with the environment and with other organizations. To overcome the problems that they face through their existence, organizations must adopt survival strategies, either individually or in group. Some questions arise at this moment:

- Can organizations be viewed as adapting individuals?
- Can organizations join together in networks and adapt as a group?
- Can this aggregation (a new form of organization) avoid survival problems?

Our aim is to answer these questions. So far we have introduced different, though compatible perspectives of the interaction between organizations and between the organizations and the environment[47]:

- an *ecological perspective*, (Sections 2.2, 2.3), derived from the work of Stinchcombe (1965), Hannan and Freeman (1977, 1984, 1989), Carroll and Hannan (2000), etc.;
- an *inter-organizational network* perspective, (Section 2.4), based on the work in the field of business-to-business marketing and purchasing: Williamson (1991), Mattson (1997), Hakansson *et al*. (1999), Ritter (1999), Gulati *et al*. (1998), Solé and Valls (1991), Ratti (1991), among others.

In addition, we have complemented these perspectives with the view from the field of *industrial economics*, (Sections 2.2 and 2.3).

The ecological perspective allowed us to face organizations as *units* in a population and to study them as a whole. The population is therefore seen as a class of organizations relatively homogeneous (the "*representative ones*") in terms of environmental susceptibility, as stated by Hannan and Freeman (1977). Consequently, some concepts

---

[47] Although the *environment* may be seen as something that is "outside" the organization, it is frequently assumed to include also other organizations, as in Agent Theory: *environment* = *landscape* + *other agents*

were introduced: *specialism*, *niche*, *legitimation*, *competition,* etc. They embody important ideas about evolution of the population and process of adaptation and selection.

The Density dependence permitted to study how a population can evolve: As Freeman (1982) observed, "*density dependence is important because it generates homeostatic processes in populations, that is, it generates equilibrium levels towards which population sizes adjust, usually at decelerating rates*".

Then, as organizations do not live isolated in their environment, we have introduced the study of inter-organizational networks. Networks incorporate a system of relationships based on a division of work and are good metaphors showing how entities in space are connected.

As we saw before, Hakansson and Snehota (1995) say that cohesion of the network is determined by certain specific forces. These forces are very important for the survival of the network and can be understood as the mechanism of the relationships that rule the network: the interdependence among agents; a structure of power; a structure of knowledge; a temporary (historic development) structure, and a spatial structure.

The issues of evolution and adaptation were referred to many times in previous chapters. Either we can see the problem of the adaptation of a group of agents as an individual characteristic of the agents – *learning* – or as a collective process bringing reproductive mechanisms into play – *evolution*. This perspective is in line with Hannan and Freeman (1977) who state that adaptive learning of individuals consists of selection among behavioural responses. Concerning organizational theory, Hannan and Freeman (1977) clearly recognize that leaders of organizations prepare strategies and so organizations adjust to the environment.

So, we maintain that the answer to the first question posed above - whether organizations can be viewed as individuals - is *yes*. That is what the ecological perspective has demonstrated. Existing members of organizational population adapt to environmental pressures by replacing less favoured competences with more favoured ones. Simultaneously, at the population level, organizational population changes, as new members holding more favoured competences compete and drive out into failure members holding less favoured competences.

Now, to answer the other two questions related to how networks adapt and survive, we need to recall some details about inter-organizational networks.

Inspirations emerging from the cross-fertilization of different research areas

As stated in Chapter 2, the study of inter-organizational networks, a recent topic in organization studies, may contribute significantly to the analysis of firm's performance. In fact, inter-organizational networks improve the chances of survival of the firms in the network, as survival is linked to firm's performance.

According to Venkatamaran and Ven (1998), the survival and growth of an organization depends on its capacity to maintain and extend its network of inter-firm relationships. But what is the impact of networking? Are networks adaptive as a group? Why do firms aggregate in the form of networks? How can the adaptation perspective be analysed in a network context? Do they *learn* collectively? How does individual learning contribute to the whole network learning?

In our view, networks can be regarded as new forms of organizations. Indeed, perhaps, they are new forms of populations, but we need to analyse more deeply some previous research to develop our proposal.

The work of Venkataraman and Ven (1998) explored the relationship between the evolution of networks and the existence of environmental shocks, as presented in Chapter 2. The authors analysed the set of relationships during the lifetime of a network and concluded that there exists a kind of "liability of newness" in network relationships.

The importance of learning in networks has been studied by Hakansson *et al*. (1999), who argued that inter-organizational networking represents *a cost-efficient way of gaining access to crucial know-how that can neither be made available internally nor be easily transferred by licensing*. One important aspect that is related to *know-how* is *innovation*. Innovation induces learning, as stated by Hakansson *et al*. (1999). Furthermore, business relationships are very important in the diffusion of acquired knowledge through a network, as presented in chapter 2.

Models of learning agents are not sufficient to answer our questions on their own. Neither do Organizational Ecology, Industrial Economics or Business-to-Business Marketing.

Organizational Ecology does not focus on the relationships between organizations, namely in the cognitive aspects; instead, it analyses organizations as a whole, in which the analysis of the characteristics of organizations and their evolution is the heart of the study.

Business-to-business marketing does not allow for a study of the evolution of populations of organizations. This approach is focused on the firm itself and on the relationships that firms maintain with the environment and with other firms. Cognitive aspects are important within this approach as they determine the behaviour of firms, activities and, consequently, what happens on the market. In mainstream economics, it is assumed that the profit maximization is the main goal of the firms.

The knowledge that is necessary for the analyses of the network adaptation and of networks benefits from the combination of the three perspectives mentioned above:

- The Organizational Ecology perspective provides a way for analysing legitimation and competition that arise with the new forms of populations;
- The Marketing perspective brings up the cognitive approach to deal with the issues of individual learning and adaptation;

- Industrial Economics offers the framework for modelling the microeconomic aspects and the market: profit, demand, supply.

Our conviction is that networks can be adaptive and that they can be viewed as new forms of populations. But we face one problem: lack of information to build the proper model and answer our research questions.

We also believe that cooperation plays a very important role in the formation of networks as it induces firms to enter in inter-organizational networks with the purpose of increasing their revenue and reducing their costs (Ebers, 1997). Therefore one of our aims is to test the effect of cooperation on the survival of firms (in an inter-organizational perspective).

As we have seen before in Chapter 2, from the technological point of view (De Woot, 1987) cooperation may be used to acquire a technology to overcome the problems of a company which has lost some of his competitiveness. As technology, knowledge and innovation can be seen as an expression of the capital of organizations, we intend to study these aspects in the scope of network modelling.

Finally, and linked to the ecological aspects of density dependence, we include space and location in our study. We know that firms adapt to birth and death of surrounding firms and properties of the space play an important role in this.

Considering the summary presented above and the research questions formulated before, we can define a set of questions (see Table 4-1).

General questions

G1: What are *the effects of organizational density* on the survival of organizations?

G2: What are *the effects of network density* on the survival of networks?

G3: What are *the effects of organizational density* on the survival of networks?

G4: How does *cooperation* motivate the formation of networks?

G5: What is the effect of individual *learning* on the formation of networks?

G6: Has the *shape of the network* some kind of influence on its performance?

G7: Does *duration* of relationships matter?

G8: What is the impact of the *distance* between the constituents of a network?

G9: How does *knowledge disseminate* through networks?

G10: How do *cognitive capabilities* of the organizations intervene in the formation and success of networks?

Table 4-1 - General questions of the research

Some of these questions have already been studied by other authors. The reader can consult Table 4-3 that summarizes for many studies their most important aspects, identify their controlled variables and the variables of the model, the main research questions and hypothesis, estimation techniques, data, and main conclusions. We also associate each of these works to the general questions listed above

However, we propose a new approach that combines different perspectives of analyses: the ecological perspective, the relational one (that deals with relationships among firms and makes the bridge with the cognitive approach), and a view from Industrial Economics, with its apparatus of microeconomic modelling.

This combined approach allows us to:

- Analyse the evolution of organizations in what concerns their adaptation to the environment and their survival; adaptation and survival can be measured by usual economic indicators, such as *profit, marginal cost*, etc.;
- Study the evolution of relationships that emerge among organizations and how these pool of relationships, and their type, can change the course of the survival of a network or a particular organization;
- Determine whether cognition (characterized under the form of beliefs, intentions and desires) plays a role in the success of organizations and networks.

In the following sections, we will use these general questions to formulate (one or more) research hypotheses. These hypotheses are tested using data coming from the empirical studies of two industry-based applications presented in Chapters 5 and 6.

4.2. General Research Methodology

The questions proposed in the previous section introduce the problem that we intend to study in this work. Data will be needed in order to support our general questions. However, there is no available information to answer all the questions posed above:

- It is possible to know how many firms are born and die every year according to the number of firms that exist in the same year. Therefore it is possible to compute the organizational density and measure its impact on actual organizational birth and death (contemporaneous density).
- Nevertheless it is very difficult to analyse the impact of the founding density (the density at the time of founding) for a particular firm. For that, we should have real time-series for which we could capture information for every firm. That way we could follow firms during their lifetime. Information should contain several time periods for the same firm and it should also cover many observations, corresponding to different firms.
- The formation of networks also requires much information. Surveys can be implemented, in which we could collect data about the type of relationships and the reasons that force firms to link to networks. Yet to analyse the evolution of

the network, several time points would be needed and therefore several interviews should be made. The shape of networks is also very dynamic and reality would be difficult to measure by the mean of surveys.

- It is also important to introduce different scenarios concerning the economic situation to analyse its impact in the evolution of firms and networks.

For the reasons presented above, we have chosen to use simulation. As explained in Chapter 3, simulation is a simplification of the world, and a well-recognized way of understanding it. It provides tools to substitute for human capabilities, as the possibility of implementing different scenarios by constructing a virtual economic world. Then, the world is populated by agents through parameter initialization and the world events are driven by agent interactions. Emergent behaviour of aggregate variables is then captured and the parameters can be reformulated in order to simulate different scenarios coming from different socio-economic perspectives.

We used computer programs to produce simulation and validated our models through initialization and comparison with available real data. Thus, we gain access to important aspects in simulation, such as control of some conditions, definition of different scenarios and prediction. We added new variables, some of them very difficult to gather, as *cognition*, *learning*, etc.

The basic methodology is similar to that we have presented in Figure 3-1 (*the logic of simulation as a method*). Our iterative methodology is illustrated in Figure 4-1.

(I) The initial conditions are defined and introduced in the simulation under the form of a set of parameters (vectors of micro-parameters, in the Fagiolo *et al*. (2005) methodology[48]).

(II) Simulation runs

(III) After the running of the simulation, the corresponding output is compared with reality (the output is a summary of some aggregate variables: the *stilyzed facts*)

[48] An example of a such a parameter is the *number of firms in the beginning of the simulation,* as defined in the Fagiolo *et al.*(2005) methodology

(IV) This comparison permits to control the evolution of the simulation and to adjust the initial parameters. The simulation can therefore be run again and results are recursively compared with reality. In this way, we can improve the quality of the model by validating the results. This iterative process ends when a predefined threshold concerning the distance between real and simulated output is reached.

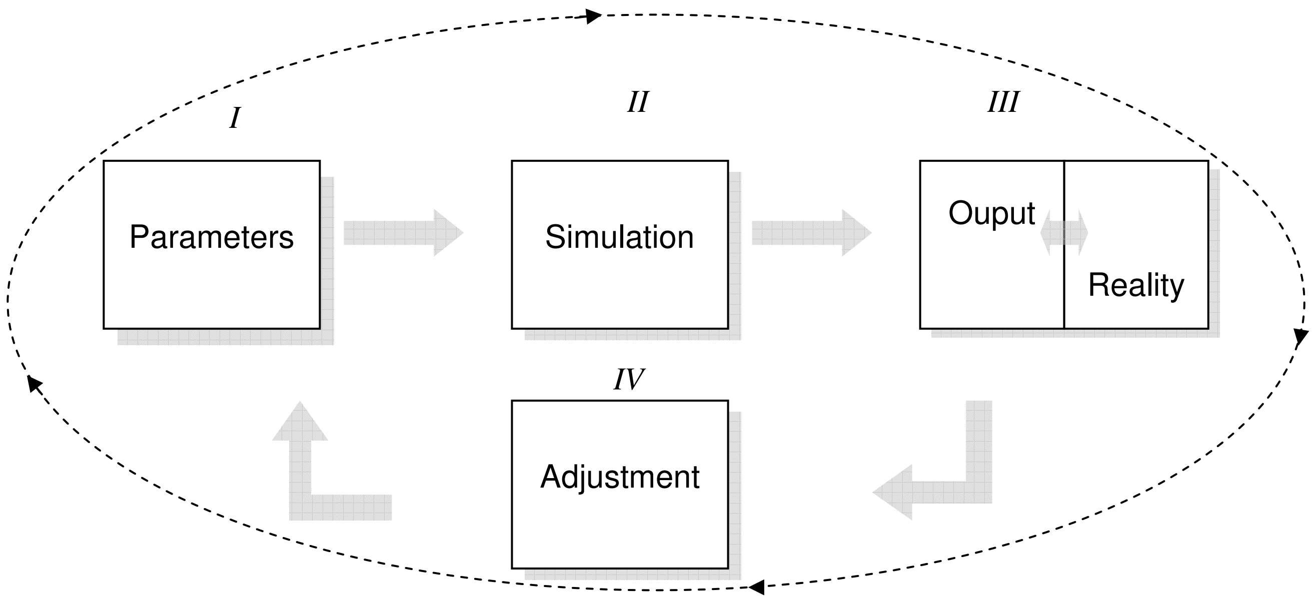


Figure 4-1 - The basic research methodology in a diagramatic form

In our work we use two main forms to adjust the parameters in order to improve the quality of the model:

a. We adjust the parameters by means of optimization algorithms like Genetic Algorithms using an appropriate metaphor of selection, crossover and mutation from Biology, based on simulated evolution (Holland, 2001 [1975]; Goldberg, 1989; Mitchell, 1997; Dumitrescu *et al*. 2000).

b. Alternatively, we can adjust the parameters using other techniques in which there is no explicit optimization for the parameters. Instead, it is up to the modeller to change the parameters in order to obtain values that seem more consistent with reality. Some of these techniques are called *historic-friendly* because they use particular historical traces in order to calibrate a model. The historic-friendly approach is strongly quantitative and mainly focuses on *microeconomic* transients of industrial paths of development, Malerba *et al* (2001).

Two studies have been developed in this work: CASOS and NetOrg. Our main objective with CASOS – Cellular Automata[49] System for Organizational Survival, presented in Chapter 5 - is to analyse the effects of certain parameters in the founding and in the mortality of organizations. A simulation, based on a Cellular Automata approach embedded with a Genetic Algorithm to calibrate the parameters permits to validate the final solution with real data coming from the Portuguese Industry.

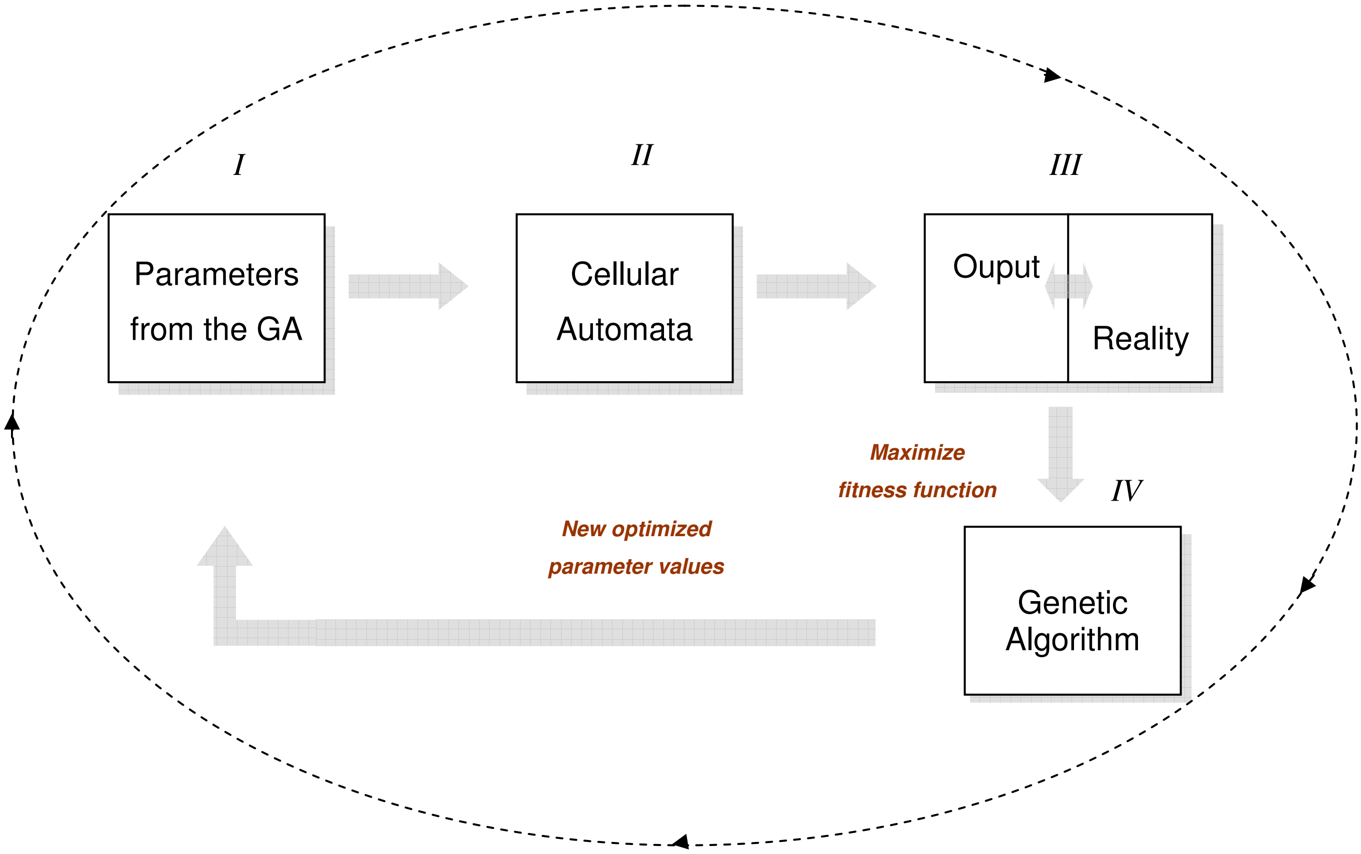


Figure 4-2 - Diagram for the research methodology focusing on cellular automata and genetic algorithms.

[49] As defined in Chapter 3, cellular automata are cells located in a regular grid where the behaviour of an individual cell is determined by a set of rules that depend on that cell and on its neighbours.

A population of hypothesis is created in which each element of the population contains codified combination of parameters represented as genes (*I*). These parameters are introduced in the simulation and the simulation runs (*II*). Then, the corresponding output is compared with reality (*III*), and (*IV*) the Genetic Algorithm searches for the hypotheses that maximize the fitness value by minimizing the distance between real and simulated output data. The whole iterative process repeats until the distance between the output of the cellular automata and reality reaches a small predefined threshold.

The second case study is NetOrg -Adaptive Networks of Organizations – a Multi-Agent framework that aims at analysing the dynamics of cooperation networks. Firms can decide to cooperate with other firms, based on cognitive and microeconomic variables. To validate our modelling approach, we have considered some evidence from the real world and decided to focus our analysis on facts that are observed in the areas of Automobile manufacturing, Textile Industry, and e-Marketplaces.

In a first stage, the parameters of the Multi-Agent System are initialized with coherent values in such a way that there is some stability in the simulation outputs. No optimization techniques are used in this case to adjust the parameters, because there is no real available data to be compared with. Therefore, the goal of this step is to tune the parameters just by analysing the admissibility of the outputs produced by the simulation. Admissibility is a somewhat subjective concept, but the idea here is to construct intervals of tolerance for the parameters in such a way that the results produced by the parameters belonging to those intervals are realistic enough to be preserved.

In a second stage, we have used these parameters to run the model for different strategies of cooperation. Here the aim is to analyse the consistency between simulation outputs and empirical studies.

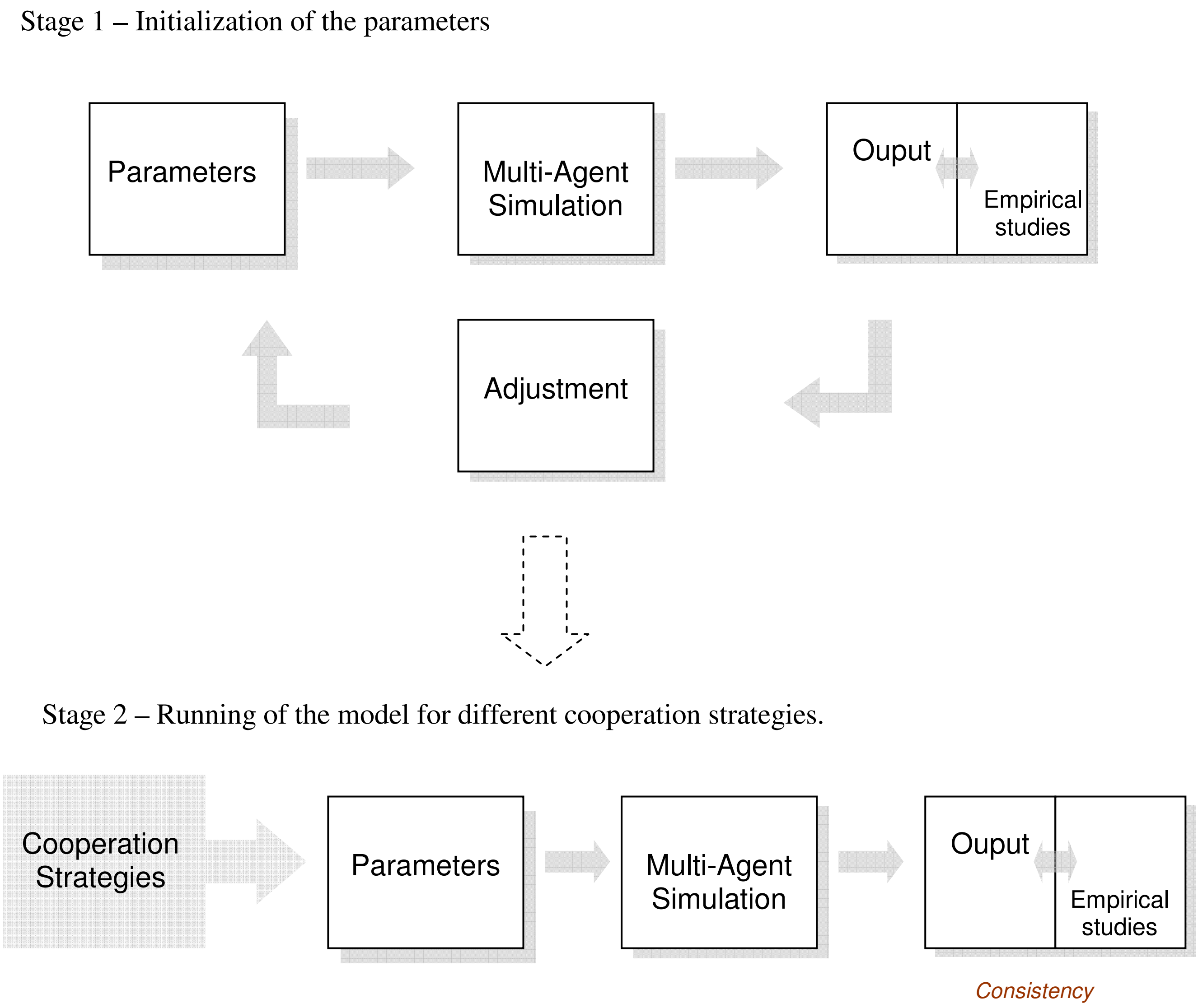


Figure 4-3 - Diagram for the research methodology using a two stage methodology.

These two case studies are developed in Chapters 5 and 6. For CASOS (Chapter 5), data from the number of existing firms as well as the number of firms' births and deaths in a particular region for a short time-series has been collected. Data needed for NetOrg (presented at the end of this Chapter) includes some personal interviews, information of cooperation in some industrial sectors and other data from the demography of firms.

The hypotheses that are tested in these two case studies are specifications of the general questions given above in Table 4-1. In this table, the association between the general questions and the specific hypotheses is established.

General question	Description	Specific hypotheses
G1	What are **the effects of organizational density** in the **survival of organizations**?	$H_{1.1}$, $H_{1.2}$, $H_{2.4}$, $H_{2.5}$
G2	What are **the effects of network density** in the **survival of networks**?	$H_{2.6}$, $H_{2.7}$
G3	What are **the effects of organizational density** on the **survival of networks**?	$H_{2.10}$, $H_{2.11}$
G4	How does **collaboration** motivate the formation of networks?	$H_{2.2}$
G5	What is the effect of individual **learning** in the formation of networks?	
G6	Does form matter? In other words, has the **shape of the network** some kind of influence in its performance?	
G7	Does **duration** of relationships matters?	$H_{2.8}$
G8	What is the impact of the **distance** in the type of relationship (vertical/ horizontal) in a network?	$H_{2.9}$
G9	How does **knowledge disseminate** through networks?	
G10	How do **cognitive capabilities** that are available in the organizations intervene in the formation and success of the networks?	

Table 4-2 – Association between the general questions and the specific hypotheses of the two case studies ($H_{1.1}$ and $H_{1.2}$. are tested in Chapter 5 using CASOS, while $H_{2.1}$ to $H_{2.9}$ are tested in Chapter 6 with NetOrg).

Some of the general questions have no corresponding specific hypothesis in the context of the studies developed here. However, statistical indicators and tests have been computed and some conclusions drawn regarding them. Besides, we define some specific hypotheses with no corresponding general questions. In such particular cases, the hypotheses are justified by the existence of the data.

4.3. Main contributions of the work

Within these possibilities of analysis, we hope to make original contributions to the way how the survival of firms has been studied. This contribution focuses in the following aspects:

- The cognitive perspective, with its characteristics (as the BDI architecture) is unexplored in the literature of Organizational Ecology. This view offers two main advantages:

    - It facilitates the individual modelling of the firm;
    - It offers a good way to understand firm behaviour;
    - It makes possible the cross fertilization of Organizational Ecology with cognitive aspects, in which the survival of firms can be viewed simultaneously as an individual and a collective issue. In fact, the collective behaviour can be analysed as the sum of the individual decisions that are modelled according to cognitive aspects.

- Examples of literature on firm survival that relates *density dependence* with the development of relationships are few or inexistent. The complexity in analysing the setting up and evolution of networks is maybe one of the aspects that make this issue difficult.

- The bottom-up analysis, typical from Multi-Agent simulation is applied to a network-based approach in an economical context. This situation is original in what concerns the use of networked MAS in an industry-based application.

4.4. Related work

Some work that is close to the general questions asked in the last paragraphs have already been introduced in previous sections. In Sections 2.2 and 3.3, the most important models related to the density dependence have been introduced in detail. In the next tables we give a summary of the main studies and their most important aspects, including the main research questions and hypothesis, and estimation techniques. A specific table concerning a summary of density dependence models using simulation techniques is provided separately.

Authors	Controlled variables	Variables of the model	Research Questions and Hypothesis	Estimation Techniques and data	Main conclusions
Wissen (2000)	Birth, death, migration and growth of business establishments.	Primary attributes of the simulation model are Age, Economic activity, size and geographical location	How can we simulate the development in size of the populations of firms, as well as its composition and spatial distribution over time? *This work is related with the general question **(G1)***	Cohort component model in a micro-simulation context called SIMFIRMS. The author makes a comparison between simulated and real data based on two official sources from The Netherlands.	Traditional demographic variables are simply not sufficient to get a hold on the dynamic behaviour of the population of firms.
Lomi and Larsen (1998)	Survival and Mortality of firms.	The model uses the density of organizations in the neighbourhood of a particular organization as the main attribute to control firm's life and death.	What are the effects of density at the time of founding on firms death, growth and decline? *This work is related with the general question **(G1)***	A Cellular Automata simulation model (similar to the Conways' Game of Life) is used to produce synthetic data. A Gompertz-Makeham function is then applied to measure organization's force of mortality. Function's parameters are estimated using the method of maximum likelihood, each parameter being associated with an effect in the model. The authors measure the impact of meaningful combinations of the founding and survival bands in estimated parameters.	Organizations founded under adverse conditions (defined in terms of relatively higher population density) have higher initial mortality rates than organizations founded in lower-density environments.
Barron (2001)	Population density and mass (total amount of activity).	Founding, failure and growth/decline as well as population density and mass which are also viewed as explanatory variables.	The goal is to compare three dynamic models of organizational populations: (i) Barnett's Model (where large organizations generate stronger competitive forces); (ii) Hannan's Model (where density becomes "decoupled" from legitimacy and competition as a population matures); (iii) Barrons' Model (where the rate of entry is not the same as the models of exit and growth). ***(G1)***	Monte-Carlo micro-simulations have been used to compare three dynamic models of organizational populations. The vital rates (founding, failure, growth/decline) have been modelled as stochastic processes within a regression model for each vital rate.	The author compares the results of these models and concludes about their complementarity (not competition) of these models, because they share common features. All three models have a variable that keeps the rate of entry low as the population ages, although density is declining, so that simulations reproduce the important empirical regularity in the trajectory of density. As a population matures, its demography seems to be increasingly determined by barriers to entry.

(cont.)

Authors	Controlled variables	Variables of the model	Research Questions and Hypothesis	Estimation Techniques and data	Main conclusions
Baum and Singh (1994)	Organizational Mortality	Independent variables - Overlap density - Nonoverlap density - Local disaggregation - Diffuse disaggregation Control variables (examples): -Organizational age -Organizational size -Profit orientation -Total number of similar firms - Demand for child care services -Number of childhood education graduates	How Organizational niches within populations influence patterns of competition and mutualism? *This work is related with the general question **(G1)***	The authors use a density dependence model focusing on the dynamics of competition (organizations diminishing each other's fates) and mutualism (organizations enhancing each other's fates) to describe the evolution of organizational populations. The following hazard function has been estimated $h_{jk}=\exp[b_{jk}X(t)]$, where h_{jk} is the instantaneous rate of transition from state j (alive) to state k (failed), X(t) is a vector of covariate values at time t (as age, size, profit orientation, total number of similar firms, demand for child care services, number of childhood education graduates) and b_{jk} is a vector of parameter estimates.	Overlap density increased mortality rates while nonoverlap density lowered mortality rates. Results also demonstrate that mutualism and competition were stronger among organizations located in the same city. "Increases in the number of population members with differential organizational niches can lead to the emergence of an organizational system comprising organizations that work together because they are different in complementary ways".
Carrol and Hannan (1989)	Organizational Mortality	-Density at the time of founding -Density at present time. - Organization age	Does density (the number of organizations in a population) affects rates of organizational founding and mortality? Is there any contemporaneous effect of density? *This work is related with the general question **(G1)***	The authors proposed a generic model for $\mu(a)$, the mortality of the organization with age **a:** $\mu(a)=\exp(\theta_1 n_a+\theta_2 n_a^2+\phi n_{f_i})$, with ($\theta_1<0; \theta_2>0; \Phi>0$). To estimate the mortality rate for each of the types of industries, a Gompertz form has been used for newspapers and breweries: $\mu_{ia}=\exp(\theta_1 n_a+\theta_2 n_q^2+\Phi_1 n_{f_i}+\Pi' X_{ia}+\Psi a)$, and a Weibull form for labour unions: $\mu_{ia}=\exp(\theta_1 n_a+\theta_2 n_q^2+\Phi_1 n_{f_i}+\Pi' X_{ia})a^{p-1}$, with a 0, 1, …, Ai. n_a denotes density at age a, and n_{fi} denotes density at time of founding of the ith organization. The authors used Maximum Likelihood techniques to estimate the model parameters.	In the formalization that Carroll and Hannan proposed, the mortality rate of the organization is modelled assuming that density at the time of founding has a positive effect on the age-specific mortality (and this effect persists over time) and that density has a nonmonotonic effect in mortality in the way that it reflects different patterns at all ages. The authors conclude that "it remains to be seen whether the delayed effect of density at time of founding is best explained by resource scarcity or niche packing".

(cont.)

-Authors	Controlled variables	Variables of the model	Research Questions and Hypothesis	Estimation Techniques and data	Main conclusions
Zhang (2003)	Formation of high-tech industrial clusters such as those in Silicon Valley	Firm's capital, Human capital, density, as well as other variables as R&D, technological innovation and imitation.	How do high-tech industrial clusters emerge in a landscape in which no firms existed before? *This work is related with the general question **(G1)***	Multi-Agent System based on a Nelson.-Winter model from evolutionary economics	The model showed that there is a contagion of entrepreneurship through peer effects that is responsible for the emergence of clusters and highlighted the importance of pioneering entrepreneurs for an emerging industrial cluster.
Mata and Portugal (1994)	Firm survival rates	Start-up size ; ownership (Number of plants); industry growth rate; entry (Log of the number of new firms in the industry); entrant's size; Industry's size; MES (Log of one half of the average size of the firms that, on average, operate 1.5 plants); proportion of industry employment in firms smaller than Minimum Efficient Scale); concentration	*This work is related with the general question **(G1)***	Duration models like the proportional Hazards Model (Con [1972]). Data were obtained from a yearly survey conducted by the Portuguese Ministry of Employment.	Larger entrants and firms that have entered with multiple establishments are more likely to stay in the market for more periods. The industries in which the expected duration of new firms is likely to be greater are those that are growing fast. A novel result is the effect of the entry rates on firm's duration which seems to be of particular importance in analysing the effect of entry on market performance.
Carayol and Roux (2003)			Is innovation a collective and interactive process that generates the formation of networks of organizations? *This work is related with the general questions **(G4** and **G9)***	Agent-Based Simulation	In this phenomenon that some call "collective invention", social interactions generate knowledge disclosure between agents belonging to competing firms which in turn stimulates innovation.
Wersching (2005)	Technological development, economic performance of firms and evolution of agglomerations in differentiated industry		Relation between innovation and agglomeration *This work is related with the general questions **(G4** and **G4)***	Agent-Based simulation	The author concluded that there is an incentive to agglomerate in young industries and that the geographical distance between them enhances innovation. Moreover, innovation and cooperation networks, and, generally, the topic of firms' dynamics have been recently studied with the help of Multi-Agent Simulation models.

(cont.)

Authors	Controlled variables	Variables of the model	Research Questions and Hypothesis	Estimation Techniques and data	Main conclusions
Purchase, and Olaru (2003)		Relationships in a B2B environment, namely Actor bonds, Activity links and resource ties	Is Centrality an important issue in network strategy? *This work is related with the general questions **(G6, G7 and G9)***	Networked Simulation	Loosing the ability to maximize opportunities due to poor network position will limit future company growth and ability to the advantages of new knowledge that may pass through the network. In order to maintain the pole position within the network, organizations must ensure that their immediate actor bonds to other organizations in pole position are strong.
Wilkinson *et al.* (2001)		Relationships between networks	What are the processes by which institutional and network structures evolve? What factors drive those processes? What is the impact of environmental conditions ? What are the ways in which the evolution of better performing institutional and network structures may be encouraged? *This work is related with the general questions **(G6** and **G7)***	NK Boolean models (developed by Stuart Kauffman)	Firms are operating in complex adaptive systems in which control is distributed through the system. No actor or entity coordinates or directs the behaviour of the network. Firms jointly create both their destiny an the destiny of others; they come to see themselves as parts of business ecosystems in which cooperative and competitive processes act to shape the dynamics and evolution of the ecosystem.
Epstein and Axtell, (1996)		Food, Birth, Death, Culture, Conflict, etc.	Is it possible to build a civilization from the bottom-up? *This work is related with the general questions **(G1, G4** and **G10)***	Agent Based simulation	*Sugarscape* is a reference Agent Based model where social structures and group behaviour arise from the interaction of individuals. Fundamental collective behaviours such as group formation, cultural transmission , combat, and trade are seen to emerge from the interaction of individual agents following a few simple rules

Table 4-3 – Summary of related works, presenting the Variables of the model, Research Questions and Hypothesis, Estimation Techniques and data and Main conclusions

Author(s)	Population Generation Method	Initialization	Procedures used for founding, death and growth	Models for vital rates	Number of iterations
Barron (2001)	Monte Carlo micro-simulation	A single "pioneer" organization is created to begin the simulation	Predicted founding, death and growth rates are estimated using empirical data by means of Weibull structure regression models	Barron compares three models (in which Ψ_t represents the rates of entry, growth or exit, depending on the signal of the involved parameters and N_t represents density at time t).	100
Lomi and Larsen (1998)	Cellular automata (maximum number of potential organizations is 8100)	Five cells are considered active in the instant t=0	Founding and death of organizations depend on the number of current organizations living in the neighbourhood (the authors consider a founding and a survival bands that they control to produce their simulated results): $\alpha^t_{i,j} = \sum_{i=i-k}^{i+k}\sum_{j=j-k}^{j+k} a^{t-1}_{i,j}$ In this expression, $\alpha^t_{i,j}$ represents the number of cells alive in a k-neighbourhood around a cell a_{ij} at time t, and $a^{t-1}_{i,j}$ defines the state of cell (i,j) at time t-1.	Transition rate (which represents a hazard or mortality rate) is a Gompertz-Makeham age-dependence model used to make the estimates comparable to those reported in the most recent empirical research)	100 (from reported examples)
Hannan *et al.* (1991)	Monte Carlo micro-simulation	A "pioneer" organization is created at the start	A predicted mean founding rate for each period is built, using an empirically estimated model of founding rates. The same applies to the death rates. Maximum Likelihood techniques have been used to estimate functions parameters	**<u>Mean founding rate:</u>** $E(Y_t) = \hat{\phi}_t =$ $= \hat{\varphi}(N_t)\exp(\hat{\gamma} N_t^2)\exp(x'_t\,\hat{\pi} + \hat{\tau}_p)$ where N_t is the density at start of the year t, $\hat{\varphi}(N_t)$ is the functional form of the legitimation function that fits best for each population (as reported in chapter 4), $x'_t\,\hat{\pi}$ contains the effects of covariates including previous foundings, and $\hat{\tau}_p$ is a set of period effects. **<u>Hazard of mortality</u>** $\hat{\mu}_i(t,f) =$ $= \exp(\hat{\theta}_1 N_t + \hat{\theta}_2 N_t^2 + \hat{\theta}_3 N_{fi})\times$ $\times \exp(x'_{it}\,\hat{\pi} + \hat{\tau}_p)\,a^{\hat{\rho}}$ where a is the age of organization and $\hat{\rho}$ is the estimate of the parameter indexing age dependence in mortality rates	150 a 300 (according to reported examples)

Table 4-4 – Summary of the main aspects of some density-dependence models that use simulation techniques

The work of Lomi and Larsen (1998) uses Cellular Automata in which founding and death of organizations depend on the number of current organizations that live in the neighbourhood. This is actually very similar to what we propose in the present work, but we are more interested in observing the impact of some covariates (that is, independent variables) in the survival of organizations.

Another relevant work is that of Rogers and Tessin (2004) which has also calibrated Agent-Based models. It is important to note that the task of parameter calibration is often made difficult by the proliferation of model parameters with non linear iterations. Rogers and Tessin use an Agent-Based model of a financial market as an example and calibrate the model using a multi-objective genetic algorithm[50].

In the next section we give empirical data of collaboration networks (to be used in Chapter 6) in three types of industries: automobile, textile, and e-Marketplaces.

4.5. The data: some empirical evidences of collaboration

To validate our model we need empirical data from the real world. In this section we present a summary of the conclusions considering the data we have analysed. The main sources of information we have consulted include official data from National Agencies, Statistical Office and Eurostat, and other sources such as industrial associations, and existing case studies.

In Chapter 5 we use real data to calibrate the simulation and therefore validate the model. Some important indicators of the Northern region of Portugal including firm births and deaths and other indicator created by the author are presented later in Chapter 5.

[50] According to those authors, the approach that consists in building models from the bottom up, allow us to concentrate more on the working of the entire system and not only on the *in* and *out* flow of data. However, this leads to a concentration of the modelling efforts on the internal mechanisms of the various agents and, thus, a proliferation of model parameters which interact in an extremely complex manner.

Concerning the collaboration networks analysed in Chapter 6, we are particularly interested in analysing the collaboration networks that emerge from the need of creating and diffusing knowledge concerning R&D activities. Recent data shows that there has been a general increase in the number of collaboration networks with R&D purposes[51].

We have decided to focus our analysis on real life examples from Automobile manufacturing, Textile Industry and Electronic markets. The reason why we have chosen these three sectors is that they constitute examples of sectors in which innovation is processed by firms in different ways. Pavitt (1984) presented a tipology of sectors according to the processing of innovation.

The Automobile sector is associated with production intensive firms with the increasing division of labour, reduction of costs, use of machinery and great dimension. Innovation aims at reducing the costs and improving the quality of the product. This sector is undergoing significant restructuring in recent decades, thanks to company reorganisation initiatives and the introduction of new technologies and organisational models. Consequently, there has been an increase in mergers and acquisitions, and the establishment of R&D partnership agreements between automotive firms[52]. In many cases these agreements take the form of networks that involve constructors (car makers) and suppliers of components.

Textile industry is a traditional supplier dominated sector of manufacturing. Supplier dominated firms make only a minor contribution to their product technology, and therefore most innovations come from suppliers. The aim of innovation is to reduce costs or to differentiate the products. In general, innovations concern the production process or the design of the product.

E-Marketplaces are a particular type of e-business, which is a term often used to identify businesses that are made through electronic means, generally the Internet.

---

[51] To support this evidence, we can state that during 2003 in Portugal, the total number of collaboration networks concerning R&D projects in the area of information technology is ten times greater than it was in 1997. In the automobile manufacturing sector, the corresponding number is four times superior for the same period (source: OCES, 2005).

[52] See EIRO – The European Industrial Relations Observatory on-line in http://www.eiro.eurofound.eu.int/2003/12/study/

Electronic inter-organizational information systems enhance the way how buyers and sellers exchange information about prices and product offerings (Oppel *et al.* 2001).

The importance of the e-Marketplaces for our study is that it is related formation of networks, although in this case the geographical distance between firms has less importance than in "traditional" business.

In the following sections, we present the main conclusions concerning collaboration networks supported by available data and put forward hypotheses that are analysed further in Chapter 6 in order to validate our simulation. A brief synthesis of the main hypotheses is presented. Details concerning the data presented in this section are described in Appendix IV.

4.5.1. Automobile Manufacturing

The work of Swaminathan, *et al.*, Selada, *et al.* (1999) and Dimara *et al.* (2003), among others, are of great importance in terms of providing good source of information concerning automobile manufacturing. The main conclusions that we can take from these works to support our hypotheses are the following (author identification and number of hypotheses to be presented and tested later on are in brackets):

- Automobile constructors promote both a concentration process of component suppliers and room for creating important collaboration networks (EMCC, 2004); (H_{21})
- Migration of constructors to emergent markets of Asia and South America as a way to rationalize production (Selada, *et al.* 1999); (H_{23})
- Contemporaneous density has a negative impact on the mortality of organizations (Carrol and Hannan, 1989); (H_{12} and H_{24})
- Density at founding has a positive impact on the mortality of organizations (Carrol and Hannan, 1989); (H_{11} and H_{25})

- If the distance between two firms is short, they have a higher probability to cooperate horizontally; if the distance between two firms is high, they have a higher probability to cooperate vertically (Dimara *et al* 2003); (H_{29}).
- Stable and long term links improve supplier's probability of survival (Swaminathan, *et al.* 2002); (H_{28})

4.5.2. Textile Industry

We gathered some case studies from the Portuguese, Spanish and Italian markets that have been made in the context of "Rede Têxtil" (CENESTAP[53], 2000a, 2000b, 2000c, 2000d, 2000e, 2001), a project supported by the Portuguese Government whose aim was to improve and promote inter-firm cooperation. In the following, we present a summary of the results obtained in those case studies that brings important details about cooperation and formation of networks.

- Networks explore the complementarity associated with the firms' specificities.
- With cooperation among firms, the time of the order satisfaction to the customers is minimized as well as production costs, and the productive capacity of the network is increased; (H_{22})
- Profitable networks indicate that one of the reasons for their success is a clear definition of the services to develop in cooperation, including, for instance, a common structure with juridical autonomy. New functions are available for some firms (as publicity and advertising) because they are accessible to all the firms in the network.
- In some cases there are signs of rivalry, when the distribution of the orders through the network is not compatible with the goals of some particular firms. There are situations of opportunistic behaviour and asymmetric benefits.

[53] CENESTAP is the Portuguese Center for Applied Textile Studies.

4.5.3. e-Marketplaces

We based our evidences in the studies of Wang and Archer (2004), Oppel *et al.* (2001), Osterle *et al.* (2001) and Camarinha-Matos and Afsarmanesh (2004). The main conclusions are the following:

- Horizontal and vertical cooperation deals with large networks.
- The links are based in long-term relationships.
- The physical distance between the elements of the network has relatively low importance.
- Transaction and search costs are reduced. The electronic search and comparison of products is one of the most important advantages of EMs. (H_{22})

These evidences will be used later in this work. In particular, the low importance of physical distance in e-Markets is considered in the definition of a specific strategy of collaboration given in Chapter 6.

In the following Chapters (5 and 6), the two applications (CASOS and NetOrg) will be presented. The modelling approach and results are considered and the research hypotheses will be defined and discussed more specifically in each of the situations.

5. Density dependence and Niche Size: a cellular automata-based approach

5.1. Introduction

The characteristics that assure the survival and growth of organizations are a topic of concern for managers and economists that search for answers to several questions such as: Why organizations get born and die? What are the determinants of their survival? What is the effect of the location in their births and deaths?

Although these questions may be answered from different perspectives, we will adopt an ecological point of view here, as space is an important component for the study of the organizational survival. Environments have spatial components that affect the evolutionary dynamics of organizational populations. Each population of organizations occupies a different space (that we call organizational niche[54]) and the size of the niche is determinant for the survival of organizations. As referred to, back in Chapter 2, a niche includes all the combinations of resources at which the population can survive and reproduce itself (Baum and Singh, 1994).

[54] More precisely, in Ecology, the total requirements of a species (resources and physical conditions) determine the space where it can live, and its success. These requirements are termed abstractly the *ecological niche*.

Due to its relevance for organizational survival, we have concentrated our analysis in the dimension of the niche. We consider that the niche dimension can be measured by its number of firms, that is to say, the organizational density[55]. Following on previous work (Lomi and Larsen, 1998), we assume that if the density value is inside the limits of a certain *survival* interval, the firm will stay alive. The same applies to the birth process (founding) of a new firm: if the density is within a *founding* interval, a new firm will be born. This approach is based on the Conway's Game of Life (Gardner, 1970) and the whole process is explained later in more detail.

To reproduce this game, we have built a cellular automata-based simulation named CASOS – *Cellular Automata System for Organizational Survival* - in which the survival and founding interval limits, as well as other variables are introduced as input parameters which we call *micro-parameters*. Then, in order to validate the quality of the simulation, the output of the cellular automata (birth and death rates of the population) is compared with birth and death rates observed in a real economy. This comparison allows the adjustment of input the micro-parameters of the simulation (the interval limits). To make this adjustment of the micro-parameters, we have used a Genetic Algorithm.

The whole process has been presented in the previous chapter as the research methodology (Figure 4-1). A generic outline of CASOS is now synthesized in the following:

(I) Definition of ***micro-parameters*** *(input parameters)*: some parameters are defined as input parameters of the cellular automata:

a. the upper and lower limits of the survival interval [DS_l ; DS_u];
b. the upper and lower limits of the founding interval [DB_l ; DB_u];
c. The *size* of the firm (S);
d. The *age* of the firm (A).

---

[55] As introduced in Chapter 2, organizational density is defined as *the number of organizations in the organizational population*.

(II) *Running of* ***cellular automata****:* depending on some rules (to be explained in the following sections), each firm can stay alive or die, and new firms can be born;

(III) ***Validation*** *(comparison between the output of the cellular automata and reality):* after running the cellular automata a certain number of iterations, we obtain birth and death rates as the output: these values are compared with real values for several periods of time obtained in empirical studies. A fitness function is built as a measure of distance between simulated and real rates.

(IV) ***Adjustment*** *of micro-parameters through a genetic algorithm*: with a genetic algorithm that uses micro-parameters as hypotheses and the fitness function, we adjust the micro-parameters to the real output. Several periods of time are considered.

The process is repeated until a threshold is attained or a pre-specified maximum number of iterations is reached.

We aim at determining the limits of the niche density interval that influence the survival and the founding of firms within. At the same time, our objective in this application is to evaluate the effects of contemporaneous and founding density on the mortality of organizations[56]

We note that there is a distinction between births and deaths in terms of the level of analysis where they are produced. Births depend on the overall density of a territorial unit, while deaths depend on the density level of a particular kind of firms.

In the next sections, we specify the modelling approach (5.2) and the technical aspects that are related to the creation of the Cellular Automata (5.3 to 5.5). Results are given in section 5.6.

---

[56] The concepts of *contemporaneous* and *founding* (or delayed) *density* were studied by Carroll and Hannan (1989), as presented in Chapter 2. The main concern of the authors was whether density at the time of founding might change the mortality rate of adult organizations. Carroll and Hannan have called this phenomenon of postponement between the cause (density at founding) and the effect (contemporary death), the *delayed density model*. They have concluded that density at time of founding had a positive effect on mortality rates in all the populations that have been studied.

5.2. The Modelling Approach

Since one of our aims in this application is to determine the density interval limits of the niche that influence the survival and the founding of firms, we need to use a set of techniques that allow us to:

- simulate the interaction between firms in a locally distributed system;
- determine the values of internal initial micro-parameters, such as density values that are determinant for the survival of organizations.

In order to achieve the first goal, we have used a cellular automata system. As we have explained in Chapter 3, cellular automata are cells located in a regular grid where the behaviour of an individual cell is determined by a set of rules which specify how the state of the cell depends on the previous state and the states of its neighbours. This is a characteristic of so called *automata* (Hopcroft and Ullman, 1979) in Computer Science. Cellular automata (CA) represent an appropriate technique when the location of the individuals is important. By using CA as a simulation technique, we expect to analyse the effects of these locally interacting automata[57] by simulating the states of the cells that live around them.

The process of adjusting the simulation micro-parameters so that some variables in the simulation would follow similar trends as real data is referred to as *calibration*. To calibrate the simulation we use a genetic algorithm. Genetic Algorithms (GA's) represent adaptive algorithms based on simulated evolution (Holland, 2001 [1975]; Goldberg, 1989; Mitchell, 1997; Dumitrescu *et al.* 2000) that use an appropriate metaphor of selection, crossover and mutation from Biology, as presented in Chapter 3.

The GA approach involves a population of hypotheses in which each element (a single hypothesis) contains a codified combination of the micro-parameters represented as *genes*. In our case, the micro-parameters are the density interval limits that we want to evaluate as the limits that are acceptable for the niche to survive and evolve. Then the

[57] When the effects of these locally interacting automata are seen as a whole, we may say that we are analysing the *emergent properties* of the system.

GA will try to search for hypotheses that maximize the fitness value, by minimizing the distance between real and simulated output data in the several periods of time.

GA's generate successor hypothesis by repeatedly mutating and recombining parts of the best current hypothesis (Mitchell, 1997), until the stopping criteria is met. In each step, every hypothesis of the current population is evaluated according to a fitness measure (Dumitrescu *et al.* 2000; Mitchell, 1997) and the hypotheses with high fitness are probabilistically selected to be the seeds for production in the next generation. A prototype of a GA is shown in Figure 5.1.

The goal of GA's is to search in a space of candidate hypotheses in order to find the hypothesis that fits best the data, that is, the one that optimizes particular criteria, evaluated by the fitness measure. The best hypothesis found by the program is given as output.

The genetic algorithm is just one of the steps in the whole iterative process that we have created in this study. As stated before, the goal of this study is to obtain the best combination of the micro-parameters (codified in the hypotheses) that characterize the size of a niche. Therefore, a population of micro-parameters is created and the cellular automata simulation starts having those micro-parameters as an input. The output of the simulation has the form of vital rates (birth and death rates) that are compared with "real" ones. The fitness function is computed as the distance between simulated and real vital rates.

Prototype of a Genetic Algorithm

Fitness: function *that is assigned to the score of a hypothesis*

Fitness_threshold*: threshold that specifies and ending criteria*

η*: number of hypotheses in the population of hypothesis H*

ρ*: proportion of the population that is going to be replaced through crossover in every iteration*

μ*: mutation rate*

γ*: number of genes that form a hypothesis*

1) Initialize the population: H←randomly generate η hypotheses ($h_1, h_2, \ldots, h_\eta$)

2) Evaluate: for each h_i in H, compute the corresponding fitness, Fitness(h_i)

3) Repeat while [$\max_h$ Fitness(h_i)]<Fitness_threshold , (i=1, 2, …, η)

Create a new generation Hs

(I) **Select**: Probabilistically select $(1-\rho)\,\eta$ members of H to add to Hs

P(h_i) is the probability for the hypothesis h_i of H to be selected:

$$P(h_i) = \frac{\text{Fitness}(h_i)}{\sum_{j=1}^{\eta} \text{Fitness}(h_i)}$$

(II) **Crossover**: Probabilistically select two members h_1 and h_2 from the $\rho\eta$ members of the population H, according to P(h_i) .

For each pair (h_1, h_2) produce two offspring applying the crossover operator. Add new members to the generation Hs.

(III) **Mutation**: select the proportion μ of members of Hs and apply the mutation. This mutation can be the swap of some of the γ genes.

(IV) **Update**: H←Hs

(V) **Evaluate**: for each member h_i of H, (i=1, 2, …, η), compute Fitness(h_i)

4) Return the most fitted hypothesis from H.

Figure 5-1– Prototype of a Genetic Algorithm - adapted from Mitchell, (1997)

Based on the fitness of each hypothesis, the genetic algorithm calibrates the solution and introduces new hypotheses in the cellular automata simulation. The process is repeated until a threshold or a maximum number of iterations is reached.

Now we take a deeper look into the details of the four steps of CASOS that were introduced in the previous section:

- – **Micro-parameters:** six parameters are defined:
    - a. the upper and lower limits of the survival interval [DSl ; DSu];
    - b. the upper and lower limits of the founding interval [DBl ; DBu];
    - c. The *size* of the firm (S);
    - d. The *age* of the firm (A)

    In terms of the Genetic Algorithm, each parameter is considered a *gene* (represented as **g**) and the whole set of six parameters is considered a hypothesis. Generically we have η hypothesis each containing γ genes. In our case we consider γ=6 and η =20.

    Formally, the first hypothesis is represented as

    $<g_{1,1}, g_{2,1}, g_{3,1}, g_{4,1}, g_{5,1}, g_{6,1}>$;

    the second hypothesis is represented as $<g_{1,2}, g_{2,2}, g_{3,2}, g_{4,2}, g_{5,2}, g_{6,2}>$;

    and then last one is $<g_{1,20}, g_{2,20}, g_{3,20}, g_{4,20}, g_{5,20}, g_{6,20}>$.

- *Running of* ***cellular automata.*** Each cell in the cellular automata represents a firm.

- ***Validation****:* we compare birth and death rates for several periods of time between simulation and real outputs. Birth and death rates are the variables to be compared. Variables are represented by V_{ij} (i=1, 2, …, n and j=1, 2, …, p), **n** being the number of time periods (t_1, t_2, …, t_n) and **p** the number of variables to be compared. Since we have only two variables (birth rate and death rate), then p=2. For each variable, we take seven periods of time, so n=7.

- ***Genetic Algorithm*** - adjustment of micro-parameters through genetic algorithm

The connection between the simulation and the calibration processes is based on the research methodology illustrated in Figure 4-1 and in Figure **4-2** and can be represented as follows:

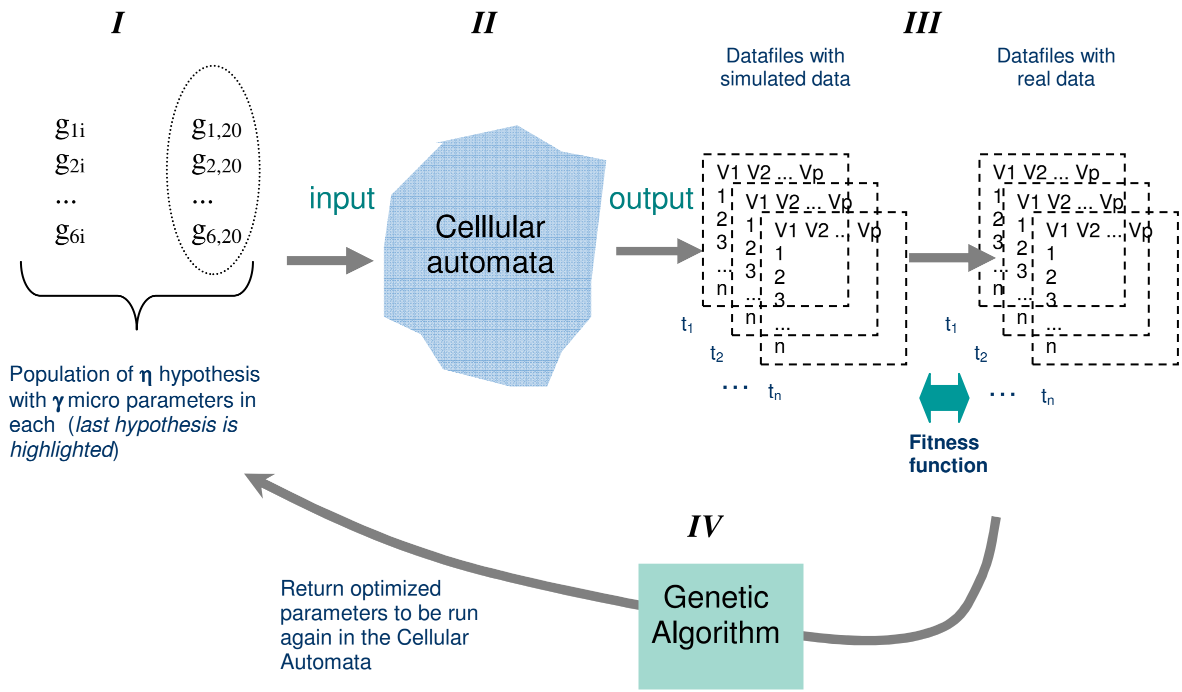


Figure 5-2- Overall architecture of the simulation with calibration of the micro-parameters[58]

The micro-parameters (*I*), represented as <$g_{1,20}$, …, $g_{6,\,20}$>, are an example of the input of the cellular automata application (*II*). Simulated outputs are then compared with real data using a fitness function that measures the distance between both (*III*). The higher the similarity between simulated and real data, the better. The task of the genetic algorithm (*IV*) is to optimize the micro parameters in order to obtain the best output considering fitness.

This process is repeated until the distance between the output of the cellular automata and reality reaches a small predefined threshold. The whole process corresponds to one run of CASOS. The main steps of CASOS are explained later in this chapter.

Next, we are going to present the components of the model that include the micro-parameters, a decisive element in this simulation.

[58] In the overall model, we consider that the population of Hypotheses of the Genetic Algorithm contains **η** hypothesis with **γ** genes in each; for the comparison of the output through the fitness function, **p** is the number of variables and **n** is the number of time periods.

5.3. Model components

Representation of the space

CASOS - Cellular Automata System for Organizational Survival - contains a landscape or space, cellular automata, and a rule. The space is represented by a graph that interconnects the places where firms can be located. We focused our environment on a particular space of four regions. Here we have chosen to study the North-West of Portugal, where textile industry is concentrated. The corresponding graph is shown in Figure 5-3. One of those regions, Ave, is often referred to as an Industrial District and is the focus of our study.

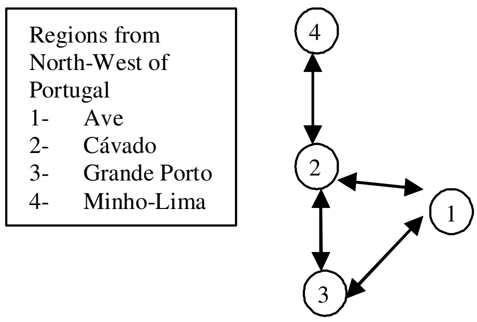


Figure 5-3 - North-western region of Portugal represented as a graph where arrows connect contiguous areas representing the four regions of the study.

Cellular automata are located in different grids, each being associated with a region, as illustrated in Figure 5-4:

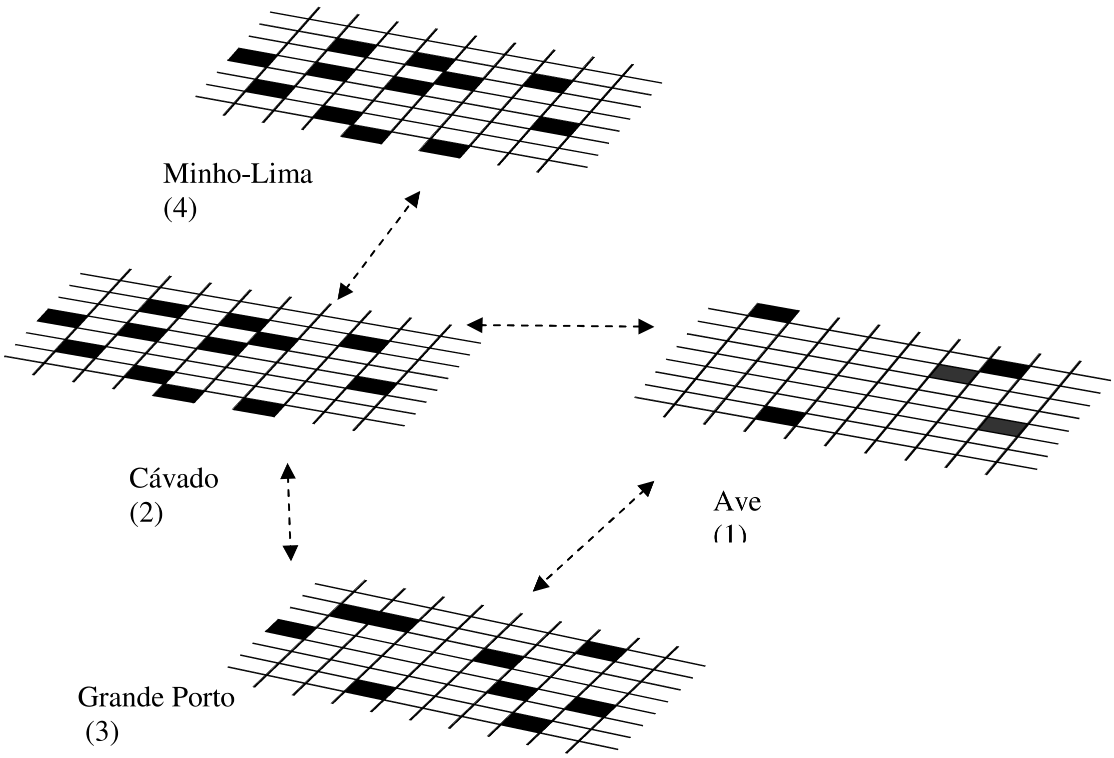


Figure 5-4 – Representation of cellular automata in the four regular grids (1 to 4) corresponding to the regions in the study. White cells in the grid represent "active " (or living) firms, while dark cells represent dead firms.

Note: The dimension of each grid is merely indicative. In practice, regions are associated with grids of different dimension.

The following table provides information about the state of the economical activity of Ave and nearby regions in 2001.

	Firms	Area	Density	Number of employers	Newborn firms	Failing firms
	Nr.	*km^2*	*nr. firms /Km^2*	*Nr*	*nr*	*Nr*
Ave	2093	1237.8	16.91	80802	164	6
Cavado	942	1242.6	7.58	28777	68	7
Grande Porto	1348	817.4	16.49	43501	75	23
Minho-Lima	130	2210.3	0.59	3886	14	4

Table 5-1 – Real data for textile industry concerning the four regions of the study relative to 2001 (Source: INE and author)

Note: the column Density was multiplied by 10

We have chosen the region of Ave as the centre of our analysis. Ave has also been frequently studied in the last decade, due to problems of competition related to emergent textile markets in Asia. Therefore, statistical information of many important economical indicators including firm creation and extinction exists and will certainly be needed to calibrate our simulation[59].

In the next paragraphs the modelling is explained with more detail, and we focus on the aspects of the cellular automata and on the rules related to the density, birth, death, size, and age.

In the cellular automata approach, each firm is represented by a *cell* with a set of attributes: location, age and size. There is one rule in CASOS that connects the cell (firm) and the environment, which is responsible for the determination of the firms' birth and death processes. To determine the birth and death of a firm, we have adopted a variant of the density dependence model.

We based our model on the basic model presented in Chapter 2, the *density dependence model*, in which vital rates of birth and death of firms are dependent on the population density. This process is modelled by the S-shaped logistic growth curve in which the growth of the population is limited to K (the carrying capacity), as seen earlier in the differential equation for population growth:

Density dependence = $\frac{dN}{dt} = rN\left(\frac{K - N}{K}\right)$**.** Equation 2-6

In this equation, N represents population size, t represents time and r defines the constant growth rate of the population. It has been proved that the solution of this differential equation is K, the carrying capacity, that is to say, the maximum size a population can attain under the conditions of the current environment. A specific interpretation of this model is that the survival of a firm depends on the number of firms that are located in its neighbourhood. We have implemented a variant of this model, using cellular automata, as explained subsequently.

[59] When no real data could be found for some periods, the data was substituted by hypothetical data. For example, some values have been estimated in Table 5.1, since there is no information for textile industry at this geographic level of analysis. Therefore, according to the weight of textile industry in the North of the Portugal (measured by the number of textile firms over the total of firms), we estimated some of the values in this table.

According to our variant, in every region, a firm can be born or die, depending on a function which is based on the number of firms that are located in its neighbourhood. For that purpose, we have calculated the density, D_j^t, involving a particular region, j, at time t, as the average number of firms per square kilometre in region j and in the nearest neighbour regions, at time t:

$$D_j^t = \sum_{i=1}^{K_j} \frac{f_i^t}{Area_i}$$ **Equation 5-1**

Density has a spatial dimension here, as it is computed as the ratio between the number of firms and the area where they are located. In this equation K_j represents the number of regions that belong to the neighbourhood[60] of region j; $Area_i$ is the area of the regions involving j, measured in square kilometres, and f_i^t corresponds to the number of firms at time t (i=1, 2, …, K_j) that are alive. The latter value is computed as follows:

$$f_i^t = \sum_{j=1}^{N_i} \delta_{ij}^t ,$$ **Equation 5-2**

where N_i represents the total number of firms in the region i and δ_{ij}^t is the state (dead - coded as 0 - or alive – coded as 1) of the j^{th} firm of the region i at time t. For example, let us consider that Ave (region 1) contains 107 "active" firms at time 2. If the neighbourhood of Ave comprises Cávado (region 2, containing 41 "active" firms), and Grande Porto (region 3, containing 87 "active" firms), then the number of firms around a particular firm in the region of Ave is given by f_1^2 = 235. The value f_1^2 is simply the sum of the number of firms that are alive in the neighbourhood of region 1. In this case we have counted all the firms of Grande Porto, Cávado and Ave.

The concept of neighbourhood is important to identify the firms that are close to a particular firm. So, we defined the *level of neighbourhood* of a particular region j, as the maximum distance that a region can be from region j, otherwise it is not considered a neighbour of region j. For instance: if the neighbourhood is 30 kilometres, then only

60 The number of regions that are considered in the neighbourhood of a particular region is computed according to the distance between regions (presented later in table Table 5-3).

regions that distance no more than 30 kilometres from region j are considered its neighbours.

In our simulation, density will affect the processes of founding and failure in the following manner: if the number of living organizations in the neighbourhood of a specific firm belongs to the survival interval [DS_l ; DS_u], then the organization will have a higher probability to stay alive. Otherwise, it will have a higher probability to die either due to “overcrowding” or to “solitude”, depending on whether the number is greater than DS_u or lower than DS_l (where DS_l and DS_u are, respectively, the lower and upper bounds of the density survival interval).

The same idea applies to the process of founding, where DB_l and DB_u are, respectively, the lower and upper bounds of the density founding interval. Moreover, considering what was said about the association between density and legitimation, DS_u may give us an idea about the maximum levels of legitimation for a particular type of industry.

We also assume that *age* and *size* affect the firm’s survival. Following Mata, Portugal, and Guimarães (1995), size is an important determinant of the chances of survival and it exerts a negative effect on the failure rate. So, we will consider that larger firms have more chances to survive. Size will be measured by the number of employees. Also, experimental studies have shown that infant firms are more exposed to death (Carroll and Hannan, 1989). Thus, we will define S_l and A_l as the lower size and age thresholds, respectively, above which an organization is more likely to survive.

Logical formulation

Therefore, as a logical formulation of what has been said before, we state that δ_{ij}^t, the state of firm i, from region j, at time t will be defined as:

$$\delta_{ij}^t = \begin{cases} 1 \text{ if } (DS_l < f_i^t < DS_u) \wedge (S_i^t < S_l) \wedge (A_i^t < A_l) \wedge \delta_{ij}^{t-1} = 1 \\ 1 \text{ if } (DB_l < f_i^t < DB_u) \wedge \delta_{ij}^{t-1} = 0 \\ 0 \text{ otherwise} \end{cases}$$ **Equation 5-3**

where S_i^t represents the size of firm i at time t, and S_l is the size threshold; A_i^t is the age of firm i at time t and A_l is the age threshold. This can be seen as the rule of the system.

Since firm size and age affect survival, they influence the state of each firm (dead or alive), which depends also on the number of neighbours (density) and on the firm's own state in the previous period.

We have introduced another variant in the original density dependence model since firms are condemned to death just if they are below the interval limits, specifically, when they are very small ($S_i^t < S_l$) or very young ($A_i^t < A_l$).

Probabilistic formulation

Therefore, assuming independence among density, size, and age, the probability that a firm remains at state 1, i.e., the survival probability for the firm i in the region j at time t, denoted $P(\delta_{ij}^t = 1)$, can be formulated as :

$$P(\delta_{ij}^t = 1) = P(\delta_{ij}^{t-1} = 1 \mid D_j^t, S_i^t, A_i^t) = P(\delta_{ij}^{t-1} = 1 \mid D_j^t) \times P(\delta_{ij}^{t-1} = 1 \mid S_i^t) \times P(\delta_{ij}^{t-1} = 1 \mid A_i^t).$$

Equation 5-4

For that reason, we have considered piece-wise linear functions normally used in fuzzy logics for each of the firm attributes - density, size, and age – for which a probability of survival was computed. Figure 5-5 shows that the probability of survival depends on the value of each of the independent variables (density, size and age).

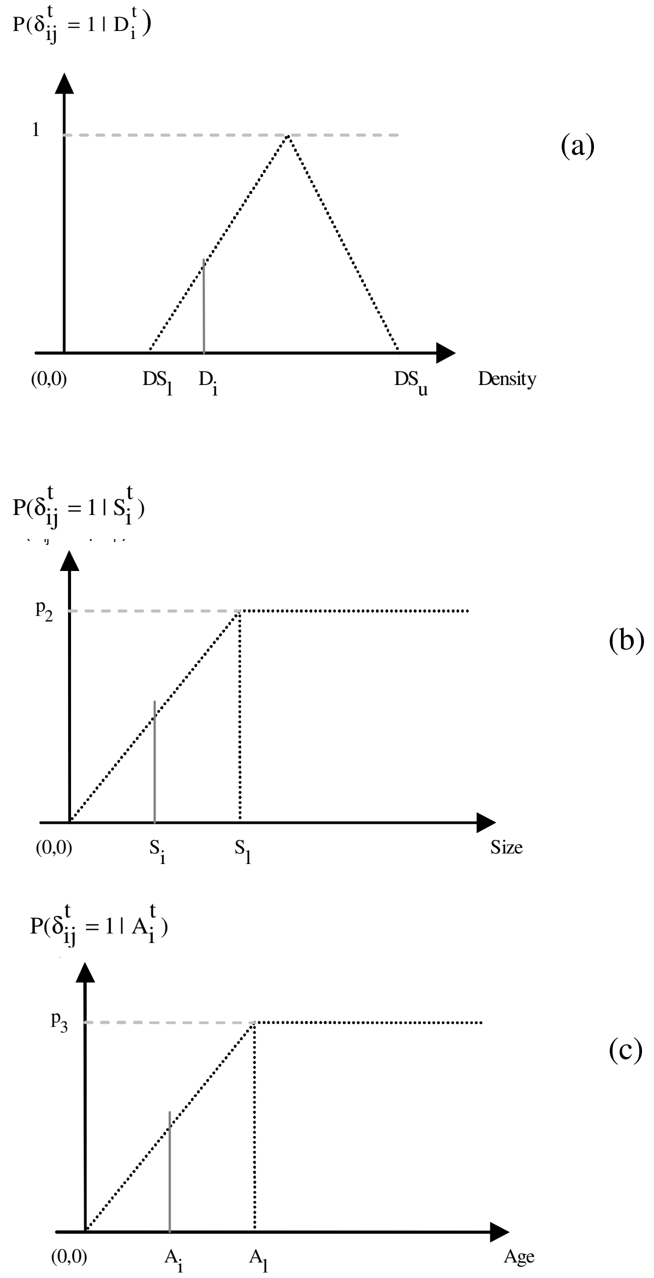


Figure 5-5: Probability of survival and interval limits for Density (a), Size (b) and Age (c) and illustrative values for D_i, S_i and A_i (at time t).

This figure illustrates how the survival probability evolves with each of the independent variables. If $DS_l<D_i<DS_u$, then, the more D_i moves towards the centre of the interval $[DS_l; DS_u]$ the higher the probability of survival (see Figure 5-5 a). The nearer S_i and A_i are to the corresponding limits for size and age, the higher the survival probability (see Figure 5-5 b) and c) respectively).

The micro parameters of the simulation (DS_l, DS_u, DB_l, DB_u, S_l, and A_l) need to be determined. They will be initialized by a random process and optimized by the genetic algorithm.

5.4. Model adjustment and evaluation

5.4.1. Generic issues concerning parameter adjustment

The calibration process

The calibration process uses a Genetic Algorithm (GA). A population of hypotheses is created in which each element of the population (hypothesis) contains the combination of parameters DS_l, DS_u, DB_l, DB_u, S_l, and A_l which are represented as genes in the hypothesis h_i, as seen earlier in this chapter. The GA then tries to search for the hypothesis that maximizes the fitness function value, that is minimize the distance between real and simulated output data.

The fitness function is associated with the measure of the distance between the values of the vital rates produced from synthetic data and real observed data. Here we have chosen to use the *Manhattan distance.* The target function is therefore a pooled Manhattan distance *d*, between two curves x_k and x_i (one for real values and the other for simulated values), spanning across the n-dimensional space - the number of compared periods - , defined by:

$$d_b(x_k, x_i) = \sum_{j=1}^{n} \left| x_{kj} - x_{ij} \right| . \qquad \textbf{Equation 5-5}$$

The Manhattan distance is similar to the Euclidean distance[61], in what concerns measuring the distance of two points in a n-dimensional space. However, it has the advantage of reducing higher differences between paired values because they are not squared differences as in Euclidean Distance.

[61] The equivalent formula of the Euclidean distance between x_k and x_i is $\sqrt{\sum_{j=1}^{n}(x_{kj} - x_{ij})^2}$

We recall that we use n=7 periods of time. Since we make two comparisons between real and simulated values (one for birth rates and another for death rates of the region of Ave for the n periods of analysis), we need to combine the two distances concerning these two types of rates. The pooled distance d (between two points, x_k and x_i, corresponding to the hypothesis h_i) is computed as given in Equation 5-6, where d_b is the distance for birth rates, and d_d the distance for death rates.

$$\varphi(h_i) = d(x_k, x_i) = \frac{d_b(x_k, x_i) + d_d(x_k, x_i)}{2}$$ **Equation 5-6**

After running CASOS for over 100 times, we noticed that the values of distance d ranged from 0 to 1000, in all cases. For that reason, we have defined that the Fitness function for the hypothesis h_i should have the form:

$$\text{Fitness}(h_i) = 1000 - \varphi(h_i).$$ **Equation 5-7**

Therefore, higher values of the fitness function correspond to lower values of the distance function.

5.4.2. Structure and specifications of the genetic algorithm

An initial hypothesis corresponding to a vector of dimension $\boldsymbol{\gamma}$ (number of involved genes or micro-parameters), taken from the population of micro-parameters, is used to initiate the cellular automata (CA) simulation. An example of such hypothesis could be the vector of values[62]:

$$\langle DS_l, DS_u, DB_l, DB_u, S_l, A_l)\rangle = \langle 2, 9, 5, 17, 1, 2\rangle.$$ **Equation 5-8**

Every time we run the CA associated with a particular hypothesis, the output values (birth and death rates) produced by the simulation are compared with real data of birth

[62] With $DS_l \leq DS_u$, and $DB_l \leq DB_u$,

and death rates. These rates are those that have been measured in the last seven years (n=7). Therefore, the fitness value associated with that hypothesis is computed using the Manhattan distance. At the end of each population of η hypotheses, the following procedure, the core of the genetic algorithm, is repeated:

1) Compute Fitness (hi)=1000-φ(hi), (for i=1, 2, …m).
2) For all population of hi (i=1,2,…,m),

   repeat while Fitness(hi) < Fitness_threshold

   a. **SELECTION**: Select the α hypotheses with the highest fitness (**α=η/2**), among the total number η of hypotheses. The second half of hypotheses (**η/2**) is generated by crossover.
   b. **CROSSOVER**: The crossover consists of finding a new population of α members (offsprings) by swapping the halves of the genes of the remaining α hypotheses. So, the proportion of the population that will be replaced through crossover, ρ is **1/2**. Every new offspring contains half of the parameters of each parent (the parents chosen randomly among the remaining α). The new set of **η/2** hypotheses is selected to the enter in the new set.
   c. **MUTATION**: Apply random mutations to a certain proportion (μ), of existing hypothesis. The process of mutation consists in randomly selecting a gene of the individual and then increase (or decrease) its value by one unit (the probability of increasing being equal to 0.5). A mechanism of preventing values from being null was implemented.
   d. **EVALUATION:** Calculate, for each hypothesis, h_i (i=1, 2, …η) the value of its fitness.
3) The best hypothesis - the one with highest value of the function Fitness(h_i) - corresponds to the smallest distance between real and simulated values and it is considered the solution of the optimization problem. The corresponding parameters that produced that solution represent the best fit for our problem. The "real" data that was used to make this comparison with simulated results is presented later on in this section.

5.5. Initializations of the simulation algorithm

Definition of the model parameters

The parameters involved in this system are the following[63]:

Parameters of the Cellular Automata	Values
Number of firms at the beginning of the simulation	1000
Steps in evolution	30
Density and ecological Interval Limits	
Survival*	lower limit (DS_l): [1; 5]; upper limit (DS_u): [6, 30]
Birth*	lower limit (DB_l): [1; 5]; upper limit (DB_u): [6, 30]
Size*	S_l: [1,5]
Age*	A_l: [1,2]
Birth Rate	20%
Neighbourhood Distance	70 Km
Parameters of the Genetic Algorithm	**Values**
N. of Hypotheses (η)	20
Iterations	10
N. of hypoth. for Selection ($\eta/2$)	10
N. of hypoth. for Crossover ($\rho=1/2$)	10
Mutation rate (μ)	10%
Fitness threshold	0,5

Table 5-2– CASOS parameters: Cellular Automata and Genetic Algorithm parameters

*Note: in the case of interval limits, one value is chosen at random from the corresponding interval

[63] The choices of the values for parameters (initial values and intervals) are based on simulation behaviour. We tested CASOS for different numbers of firms and generations, the results not being surprisingly different. The only drawback was the duration of the runs, which could take more than three hours, if a great number of firms (more than 2000) or generations (more than 50) were considered. It was difficult to deal with such a situation, as we needed to repeat the simulation for different runs in order to test the results. So we took 1000 firms and 30 generations as initial values for these parameters. Concerning density and ecological parameters, as well as the parameters of the genetic algorithm, our choice was based in large intervals (in order to allow parameters to be chosen at random) and in the coherence of the values involved.

The initial location of firms in the beginning of the simulation is made according to an *attraction level* defined later on in this section. The interval limits of Survival, Birth, Size and Age were defined in such a way that the output results could show some coherence. Values outside those intervals would produce irregularities on the simulation (as total extinction of firms).

Although the birth rate is set to 20%, the death rate is not predefined explicitly in this simulation. The reason for that to happen is related to the way how birth and death processes are modelled in Organizational Ecology literature. As explained above, births depend on the overall density of a territorial unit. So, a fixed rate of 20% is multiplied by the existing firms in that territorial unit to obtain the number of newborn firms.

On the other hand, the death process is assumed to be related to every individual firm. It depends on the density level of a particular kind of firms. Therefore, the death process is modelled according to the probability of survival presented Equation 5-4, that is, the survival of individual depends on the values of Density, Age and Size.

The Data

We have introduced real data in the simulation. A distance matrix was defined so that it would be possible to compute the number of neighbours in every region (the neighbours are those regions that share contiguous borders). The distances between the main cities of each region are given in Table 5-3.

	Ave	Cávado	G. Porto	M. Lima
Ave	0	35	70	50
Cavado	35	0	50	40
G. Porto	70	50	0	75
M. Lima	50	40	75	0

Table 5-3 – Distance matrix between the regions of the study

To compute the density at every instant t, D_j^t, involving a particular region, j, we need the area of every neighbour region, as shown in Equation 5.2. Therefore, we have considered the column "Area" from Table 5-1 that contains the areas, in square kilometres, for the four regions in the study.

Following Hannan and Carroll (1992) we assume that some regions are more favourable (more attractive) to founding. These are the regions that have, for instance, more roads and transportation conditions as these enable the flow of resources. Therefore, we have also defined a level of attractiveness for each of the regions in the study. This attractiveness is the consequence of externalities[64] that emerge in some industrial clusters and are related to the space and to the niche. Hannan and Carroll (1992, pp. 98) refer to this as the "*effect of interaction between subpopulations*". They suggest that the effects of growth in some kind of populations improve the life chances for other populations by stimulating, for instance, the flow resources into a system. In fact, they refer that "*interdependence among subpopulations has considerable interest and importance when their fundamental niches intersect substantially*".

Consequently, we have based our estimate of the region attractiveness in the combination of some variables as: education level, number of roads, population density, etc. Table 5-4 provides the *attraction level* that was computed for each one of the regions[65]:

[64] An externality is an effect from one activity which has consequences for another activity but is not reflected in market prices. An externality occurs, for instance, when a firm builds a new road in order to transport its products faster and this decision causes benefits to other firms in the neighbourhood. Externalities can be either positive, when an external benefit is generated, or negative, when an external cost is generated.

[65] The attraction level was computed in the following manner: for each region, a rank between 1 and 10 was defined, taking into account the price of land (PRICE), the number of roads (ROADS), the educational level (EDUC) and population density (POP). The attraction level was defined as a function, with the form: 0.2xPRICE+0.3xROADS+0.3xEDUC+0.2xPOP. After obtaining a real value for each region, an integer rank was defined (between 1 and 10) and assigned to the corresponding region. Then, the attraction percentage of a region i was computed as the proportion between the attraction levels of the region i and the sum of all the attraction levels of all the regions.

	Attraction level	Attraction percentage
Ave	7	0.41
Cávado	4	0.24
Grande Porto	5	0.29
Minho-Lima	1	0.06

Table 5-4– Attraction level of the regions in the study.

After computing a pooled combination of the variables, we built an index of attractiveness for every region and converted it into a percentage (*attraction percentage*). Then we made a correspondence between this attraction percentage and a level, the *attraction level*. Therefore, firms' initial location is made according to this level. For example, in a population of 200 firms we expect to have 82 firms in Ave, 48 in Cávado, 58 in Grande Porto and 12 in Minho-Lima.

Finally, to make proper comparisons between "real" and simulated data, we have collected the birth and death rates for the firms of the region of Ave in the last seven years (the number of time periods was defined before n=7). The following table contains this data:

	1997	1998	1999	2000	2001	2002	2003
Birth Rate	78.36	92.71	70.05	59.14	63.74	84.38	44.95
Death Rate	2.87	10.66	4.04	3.64	3.26	5.18	10.67

Table 5-5– Evolution of real birth and death rates of firms in the region of Ave ($\%_0$)

Source: data is based on INE (2002 to 2004)

The fitness function compares simulated data with this real data. Since a distance-type measure was used (Manhattan distance), lower values in the distance function correspond to situations where the simulated values are closer to the real values. As an illustration, let us assume that the GA returned the following best hypothesis:

$$\langle DS_l, DS_u, DB_l, DB_u, S_l, A_l \rangle = \langle 2, 15, 5, 24, 1, 2 \rangle. \qquad \textbf{Equation 5-9}$$

In CASOS, after running the simulation, this set of parameters produced the following simulated birth and death rates:

	1997	1998	1999	2000	2001	2002	2003
Birth Rate	68.81	86.54	121.75	25.27	75.16	73.48	17.87
Death Rate	51.05	70.80	53.17	21.83	17.88	6.89	12.41

Table 5-6– example of values for birth and death rates of firms produced by CASOS

Therefore, according to what has been said before, the value of the fitness function is computed combining the Manhattan distance between the real and simulated values for birth and death rates. In this example, the distance for "births" is computed as follows:

$$d_b(x_k, x_i) = \sum_{j=1}^{p} \left| x_{kj} - x_{ij} \right| =$$
$$= |78{,}6 - 68{,}81| + |92{,}71 - 86{,}54| + |70{,}05 - 121{,}75| + |59{,}14 - 25{,}27| +$$
$$+ |63{,}74 - 75{,}16| + |84{,}38 - 73{,}48| + |44{,}95 - 17{,}87| = 150{,}71$$

Equation 5-10

The Manhattan distance for death rates is computed in the same way:

$$d_d(x_k, x_i) = \sum_{j=1}^{p} \left| x_{kj} - x_{ij} \right| =$$
$$= |2{,}87 - 51{,}05| + |10{,}66 - 70{,}80| + |4{,}04 - 53{,}17| + |3{,}64 - 21{,}83| +$$
$$+ |3{,}26 - 17{,}88| + |5{,}18 - 6{,}89| + |10{,}67 - 12{,}41| = 193{,}72$$

Equation 5-11

The final pooled distance is the simple average of these distances:

$$\varphi(hi) = d(x_k, x_i) = \frac{d_b(x_k, x_i) + d_d(x_k, x_i)}{2} = \frac{150{,}71 + 193{,}72}{2} = 172{,}22$$

Equation 5-12

Hypotheses to be tested

The hypotheses of the GA that, to some extent, validate the model presented in this section reproduce the effects of density at founding and contemporaneous density on the mortality of organizations. As it was analysed in Section 2.3, density at founding has substantial implications on the survival of organizations while contemporaneous density (the number of firms existing at the moment of the death) has a corresponding negative effect. The effect of the size on firm survival is also negative.

Therefore, in the Table 5-7 we have identified three hypotheses to be tested with survival analysis and the corresponding links (if they exist) with the questions proposed in Chapter 4.

Specific Hypothesis	Empirical evidence	Methodology	Correspondence with general questions
H11	**Density at the time of a firm founding** has a positive impact on the mortality of organizations	Survival Analysis	G1
H12	**Contemporaneous density** has a negative effect on the mortality of organizations	Survival Analysis	G1
H13	**Size** has a negative effect on the mortality of organizations	Survival Analysis	

Table 5-7 - Hypotheses to be tested in CASOS.

Main steps of the simulation

As we saw earlier in Figure 5-2, the CASOS simulation algorithm uses four steps (numbered *I* to *IV*). Now, we give the pseudo-code version of CASOS making the proper correspondence with the four steps presented back in Figure 5-2. As CASOS runs in cycles, we must distinguish the first iteration from the following. Each run of CASOS corresponds to ten iterations of twenty hypotheses each.

Pseudo-code of CASOS

- ***In the first iteration***

- **Start-up***: define a fixed number of initial firms that are randomly distributed among the cells of the different grids in the regions, taking into consideration the attraction level of each region. Those cells are considered active (or alive) firms at the instant t=0.*

- *(I)* **Define the micro-parameters** of the model (*upper and lower boundaries for survival and founding, and lowe*r limits for age and size). In the first run of CASOS these parameters are defined randomly within their corresponding interval limits. Each set of micro-parameters corresponds to a hypothesis for the GA.

- ***Repeat for all iterations***

- Repeat the micro-parameter adjustment process using GA and evolution. *Repeat until a predetermined threshold for the best fitted hypothesis has been reached or a pre-specified number of iterations has been attained, whichever happens first.*

- *(II)Repeat the evolution of firms (using the parameters of a specific hypothesis); repeat (II) as many times as the number of different hypotheses (since each hypothesis contains a combination of the micro-parameters to be tested)*

  - Produce details for N periods of time; Determine in which 'regions' a new firm will be born; Determine which firms will die; Firms that survive can grow (increment their size);

  - (III) Compare vital rates (birth, death) of the population of firms based on the simulated data (at the region aggregate level) with those from real data and calculate the fitness function value to obtain the best hypothesis (using the distances discussed later in this section).

- *(IV) Apply Selection, Crossover and Mutation for the set of* η *hypotheses, and produce a new generation of* η *hypotheses.*

- ***Return the best hypothesis***

Figure 5-6 – Pseudo-code of CASOS.

5.6. Experimental results

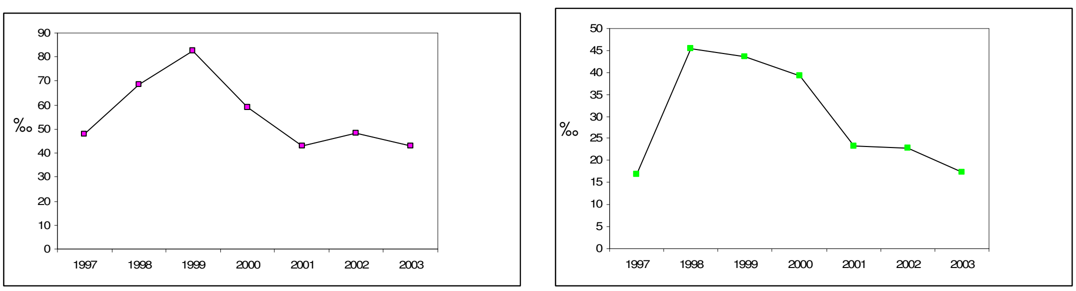
5.6.1. Evolution of the data

After running CASOS for more than hundred times (each run of CASOS corresponding to ten iterations of twenty hypotheses each), a final set of parameters is selected, constituting the optimized solution (i.e. a hypothesis) for the problem. However, there does not seem to be a unique solution, since different runs of CASOS produce slight variations in the final solution. The next table contains five solutions or hypotheses (identified in the first column) for different runs of CASOS. For every solution, the corresponding distance was computed.

h_i	DS_l	DS_u	DB_l	DB_u	S_l	A_l	Distance $\varphi(h_i)$	Fitness (h_i)= $=1000-\varphi(h_i)$
1	4	9	4	24	3	2	136.64	863.36
2	3	15	1	23	2	2	148.7	851.3
3	3	10	4	25	3	1	153.04	846.96
4	3	11	3	25	3	2	163.67	836.33
5	4	12	5	24	4	2	197.03	802.97

Table 5-8 – Simulated solutions of the calibration process in CASOS: density interval limits (DS and DB, multiplied by 10) and Age and Size limits.

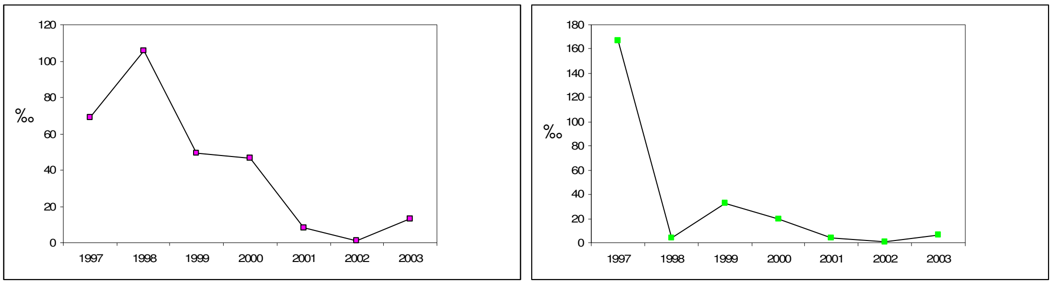
Each solution corresponds to a value of the distance function φ (the pooled Manhattan distance) which establishes a value of the proximity between the simulated and the real death and birth rates. The following figures give an idea about how different solutions presented in this table correspond to different outputs, i.e. death and birth rates.



(a) (b)

Figure 5-7– Simulated birth (a) and death rates (b) from the solution h_1:

h_1= <DS_l, DS_u, DB_l, DB_u, S_l, A_l> = <4, 9, 4, 24, 3, 2>

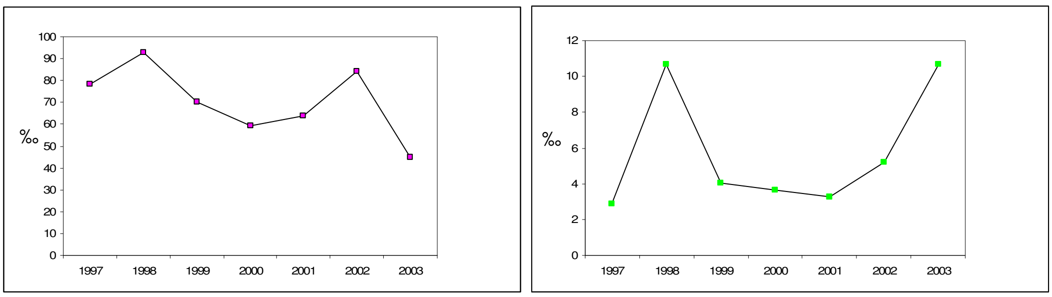


(a) (b)

Figure 5-8 – Simulated birth (a) and death rates (b) from the solution h_6:
h_6=<DS_l, DS_u, DB_l, DB_u, S_l, A_l>=<4, 10, 3, 23, 4, 2>

The two solutions presented here h_1=<4, 9, 4, 24, 3, 2> and h_6=<4, 10, 3, 23, 4, 2> are quite similar, but lead to differences in terms of the birth and death rates that they have produced in the simulation. Their distance from real values is also different: h_1 has a higher fitness value, Fitness(h_1)=863,36, while the fitness value for h_6 is much lower, Fitness(h_6)= 774,05 (the maximum of the fitness is 1000).

To help understanding the value of the fitness associated to the solutions represented in these figures, let us depict the real values of the birth and death rates that were presented in Table 5-5.



(a) (b)

Figure 5-9– Real birth (a) and death rates (b) of Ave between 1997 and 2003 (source: INE)

The shape of the birth rate for the firms in Ave shows two peaks during this six-year period, but it seems to be decreasing after 2002. The death rate, which is very much lower than the birth rate, attained a top in 1998, and it has been increasing strongly after 2001.

As we can see from the analysis of Figure 5-7, Figure 5-8, and Figure 5-9, the value of the fitness can be used as a measure of comparison between real and simulated rates. Besides, for more similar shapes, the fitness tends to increase. However, the Manhattan distance does not take into account the order of the values (that are associated with the years) of the vectors being compared.

Modified Method of Computing Distances using Growth Distances

We have decided to consider the order of the values (real and simulated) with the aim of validating the final fitness value. Only the region of Ave has been considered in order to get more focused view of this problem. We created a new distance function to compare the corresponding values of the simulated and real growth. In this way it is possible to compare the evolution of the shapes between the real and simulated curves. For each period of time, t (each period corresponding to one year), a difference Δt (with t=1, 2,…, 7) has been computed separately for death and birth rates at instants t and t+i (i=1, …6), as shown in Figure 5-10**.**

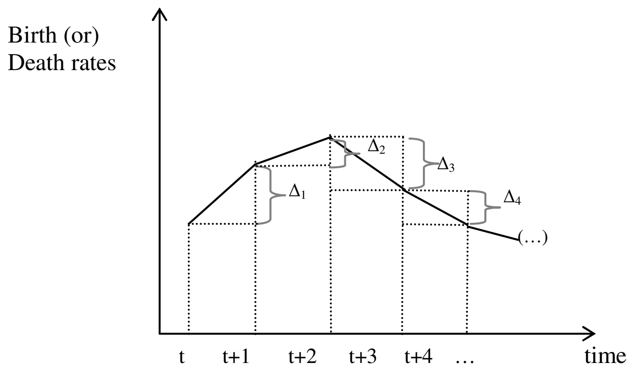


Figure 5-10 – Representation of the differences of consecutive values in the evolution of birth or death rate.

In this figure, Δ_1 represents the difference of the considered rate (birth or death rate) between instants t and t+1; Δ_2 represents the difference between instants t+1 and t+2, etc. Then, we used a distance to compare these differences in real and simulated values. This *growth* distance is described in equation 5.13:

$$d_{\Delta}(x_k, x_i) = \sum_{j=1}^{n-1} \left(\Delta_{kj} - \Delta_{ij}\right)^2 ,$$ **Equation 5-13**

where Δ_k and Δ_i represent two curves of dimension 6 (one with real and other with simulated values) each containing the differences explained above. We use this function to compute the distance between the curves x_k and x_i, through the Euclidean distances of the correspondent differences of consecutive values. The reason to use the Euclidean distance is that we respect the order of each of the values in the curves x_k and $x_{i.}$ As we did before (see Equation 5.6), we compute two distances corresponding to Equation 5.13: one for the birth rate and the other for the death rate. The final pooled distance is defined by:

$$d_{\Delta}(x_k, x_i) = \frac{d_{\Delta b}(x_k, x_i) + d_{\Delta d}(x_k, x_i)}{2},$$ **Equation 5-14**

in which $d_{\Delta b}$ represents the *growth* distance between birth rates (real e simulated) and $d_{\Delta d}$ represents the growth distance between death rates (real e simulated). To complement the analysis we have also calculated the *Euclidean* distance between the curves:

$$d_e(x_k, x_i) = \sqrt{\sum_{j=1}^{n} \left(x_{kj} - x_{ij}\right)^2}$$ **Equation 5-15**

We also considered a pooled distance (combining the birth and death rates), as in equation 5.14. We used the new target function with the *Growth* distance as fitness function[66] and ran CASOS again. The following table presents a comparison, for a particular run, between the values of the fitness obtained with the *Growth* distance functions and the other two (*Manhattan*, and *Euclidean*)[67]:

---

[66] The corresponding fitness function has been computed as in equation 5.7. Fitness(h_i)=1000-φ(h_i), where φ(hi) is the the pooled distance similar to equation 5.14

[67] We recall that that the fitness of the hypothesis h_i is obtained from the distance φ(h_i), as follows: Fitness(h_i) = 1000-φ(h_i)

#	Growth	Euclidean	Manhattan
(1)	927.0	902.6	774.9
(2)	956.2	880.5	784.5
(3)	942.2	911.2	774.8

Table 5-9 - Comparison between the fitnesses of the Manhattan, Growth, and Euclidean distances in three different solutions in which the Growth distance has been used as Fitness function.

The values presented in Table 5-9 are the final fitness corresponding to particular solutions[68] of CASOS. During the simulation process several values of fitness are produced in each iteration, as presented before in Figure 5-6 *(step III),* and the optimization process (via GA) leads to the one that minimizes the distance between real and simulated data.

Figure 5-11 illustrates the evolution of the simulated birth rates corresponding to the fitness values of the Growth distance presented in Table 5-9.

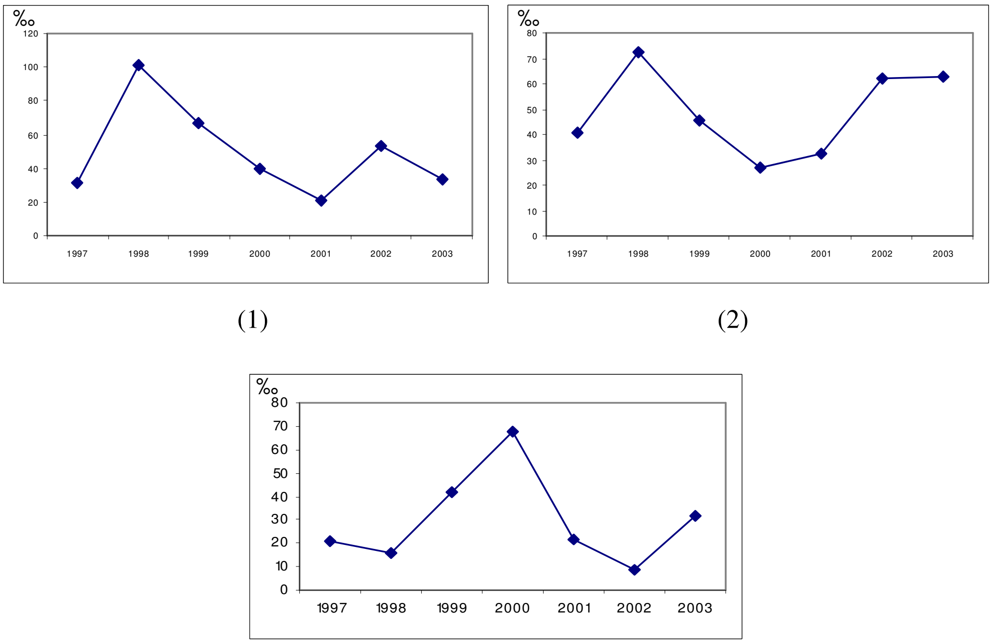


Figure 5-11 – Evolution of the simulated birth rates of Ave corresponding to the Fitness values of the Growth distance presented in table 5.9

[68] These solutions are the final of particular runs of CASOS, but they constitute a set of illustrative examples, the best solutions being presented later in this section (we needed to run CASOS for many times to obtain higher fitness). The solutions that produce the fitness presented in Table 5-9 are the following:

- $<DS_l, DS_u, DB_l, DB_u, S_l, A_l>=<3,13,3,16,3,1>$;
- $<DS_l, DS_u, DB_l, DB_u, S_l, A_l>=<5,24,4,14,1,1>$;
- $<DS_l, DS_u, DB_l, DB_u, S_l, A_l>=<3,13,3,16,3,1>$

Comparing these graphs with the graph of Figure 5.9 (a), we can observe that these shapes are more similar to real values than those in Figure 5-8. Nevertheless, we can observe other differences (in scale), that arise because we are comparing the shapes. In fact, we are comparing the differences (growths) in every step of evolution in both curves and do not take into account the values themselves. Consequently we decided to combine those functions by using a measure that takes in consideration both the evolution of the shapes and the values.

Combining the Two Approaches of Calculating Distances

In order to improve the form of the fitness function we decided to combine the two approaches. One possible key for the problem was to pool the distances taking into consideration the relationships between them. Therefore, we obtained the distances (Growth, Manhattan and Euclidean) for different runs of CASOS and calculated the Pearson correlations between them (Table 5-10).

	Growth	*Euclidean*	*Manhattan*
Growth	1		
Euclidean	-0.09724	1	
Manhattan	-0.2548	0.776442	1

Table 5-10 – Spearman correlation coefficients between the three types of distances used in the curve's comparison

We observe that there is almost no linear association between the Growth and Euclidean distances, since the absolute value of the correspoding correlation (0.097) is very small. It is interesting to notice that the relationship between Growth and Manhattan is -0.25. For those reasons, we decide to create a new *combined distance* by pooling the Growth and Euclidean distances (conferring a higher weight to the growth distance), as follows:

$$d_{comb} = 0.6\ d_{\Delta} + 0.4\ d_e \qquad \textbf{Equation 5-16}$$

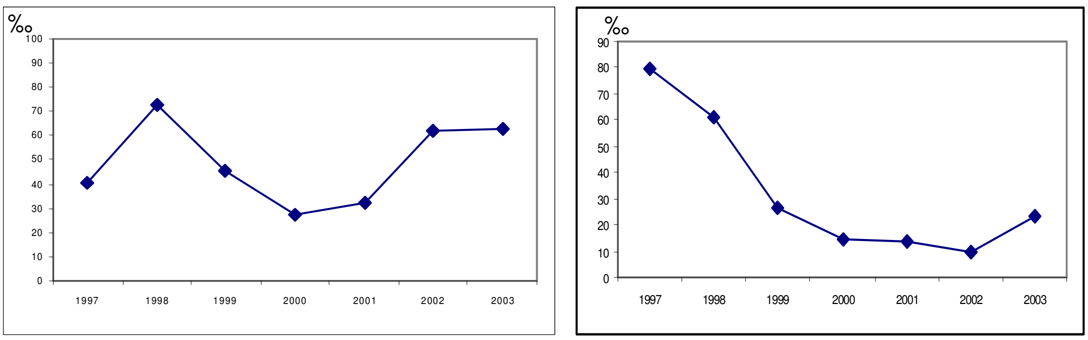
where d_{comb} is the new combined distance function that will be used as the fitness function in CASOS algorithm, d_{Δ} represents the Growth distance and d_e represents the Euclidean distance. Next, we present the best results of the comparisons between real and simulated values using this combined distance function:

h_i	Fitness(h_i) = 1000-φ(h_i)				Solutions					
	Growth	Euclidean	Manhattan	Combined	DS_l	DS_u	DB_l	DB_u	S_l	A_l
1	***943.2***	***922.1***	***817.9***	***934.8***	***5***	***28***	***2***	***20***	***2***	***1***
2	942.2	911.2	774.8	929.8	3	13	3	16	3	1
3	943.2	900.4	761.7	926.1	5	28	2	20	2	1
4	956.2	880.5	784.5	925.9	5	24	4	14	1	1
5	950.9	886.7	814.5	925.2	4	10	3	25	4	1

Table 5-11 - Comparison between the fitnesses of the Manhattan, Growth, Euclidean and Combined distances of five solutions h_i, (i=1, 2, …5) in which the Combined distance has been used as the Fitness function.

We note some heterogeneity in the solutions, namely in the values of DS_u. The best fitness we obtained using the Combined fitness function was 934.8, corresponding to the solution:

$$\langle DS_l, DS_u, DB_l, DB_u, S_l, A_l\rangle = \langle 5, 28, 2, 20, 2, 1\rangle \quad \text{Equation 5-17}$$



(a) (b)

Figure 5-12 - Simulated birth (a) and death rates (b) from the solution h_6:

$h_6 = \langle DS_l, DS_u, DB_l, DB_u, S_l, A_l\rangle = \langle 5, 28, 2, 20, 2, 1\rangle$

The evolution of simulated birth rates is similar to the evolution of the real ones (see Figure 5-9). However, simulated and real death rates show different evolutionary paths.

Comparison Between Simulated and Real Data

We used a Wilcoxon test for paired samples (Conover, 1999) in order to compare simulated birth and death rates with real ones. Starting with birth rates, we considered that the set of hypotheses to be tested is:

H_0: $\mu_s = \mu_r$

H_1: $\mu_s \neq \mu_r$,

in which μ_s and μ_r represent, respectively, the simulated and real birth rates means[69]. We concluded that the hypothesis H_0 tends not to be rejected (p-value>20%), confirming the similarity between and real birth rates[70]. Considering the death rates, we tested an identical set of hypothesis (in which μ_s and μ_r represent now, respectively, the simulated and real death rates), but the similarity between simulated and real paths was not confirmed since the corresponding hypothesis H_0 was rejected (p-value<1%).

We found a relationship between simulated and real curves in the majority of the solutions, concerning the birth rates. Besides, the differences in the absolute number of firms per region are similar if we compare real data with simulated data: Ave registers the highest number of textile firms, and Minho-Lima the lowest[71].

Although the similarity between the real and simulated rates may represent a validation of the simulation model in a certain way (and can be used as a validation form in the context of the evolutionary path of industry in these regions as we have presented in Section 3.3.3), we are more interested in analyzing further aspects of the survival of organizations that can be estimated from the simulation, rather than searching for a

[69] The Wilcoxon Signed Ranks test aims at studying the position or ranking of data from two related samples (defined by 1 and 2). The null hypothesis (H_o) for this type of test is: $\mu_1=\mu_2$, which states that there is no difference between the samples 1 and 2, μ_1 and μ_2, representing respectively the medians (or means in the case where the samples are both symmetric) of the samples 1 and 2. Usually, the alternative hypothesis H_1 is bilateral ($\mu_1 \neq \mu_2$) meaning that there is a difference the two samples.

[70] We used the same test to analyse the similarity between birth rates (simulated and real ones) considering the 22 most fit solutions, and did not reject the hypothesis H_0 in fourteen of them.

[71] In order to test if real and simulated data were similar in what concerns the total number of firms, we used a Wilcoxon Signed Ranks Test for paired samples in the four regions of the study. Considering that the hypotheses to be tested are H_0: $\mu_i = \mu_j$ *versus* H_1: $\mu_i \neq \mu_j$, in which μ_i and μ_j represent, respectively, the simulated and real means of the absolute number of firms in each region (i and j, with $i \neq j$), we concluded that the hypothesis H_0 was not rejected for Minho-Lima and Cávado (p-value>5%) and was rejected for Ave and Grande Porto (p-value<1%), meaning that the similarity between the paths is significant in the latter case.

perfect fit between simulation and reality. In any case we need to be aware of the problem of overfitting the data.

In the next section, we perform the survival analysis to measure the impact of other variables in the survival of the firms.

5.6.2. Survival Analysis

To validate some properties[72] of the model and to measure the impact of the variables in the survival of the firms, we have used a technique called *survival analysis*. Generally, *survival analysis* is seen as a collection of statistical methods where the outcome variable is the *time until an event occurs* (*time-till-event*). Like in linear regression, there is a model of the relationship between predictor variables and time-till-the-event. One of the reasons to use a survival analysis technique such as *Cox regression* (Cox and Oakes, 1984), instead of linear regression is that we may have the so called *censored data* in our sample, that is, some of the cases may have not have resulted in the terminal event (i.e., firms may be not *dead* at the end of simulation)[73].

The Cox Regression is useful for modelling the time to a specified event, based upon the values of given predictors (also called *covariates*). The basic model offered by the Cox Regression is the proportional hazards (PH) model (Cox and Oakes, 1984), a very popular model in survival analysis. One of the reasons why the PH model is so popular is that the distribution of the survival time variable need not be specified. However, if it is believed that survival time follows a particular distribution, then that information can be used in a parametric modelling of survival data, as the Weibull model (Kleinbaum and Klein, 2005).

[72] The validation consists in comparing the effects of independent variables obtained by survival analysis with the corresponding effects described in existing literature.

[73] Survival Analysis (or Failure Time Analysis) was primarily developed in the medical and biological sciences, although it is also widely used in the social and economic sciences, as well as in engineering (reliability and failure time analysis). In economics we may study the "survival" of new businesses. In this case, a censored observation may occur if a firm is still alive by the end of the study period.

In our case, as we do not assign any distribution to the survival time, we will use the PH nonparametric model. The PH model assumes that the time to event and the covariates (or predictors) are related through the following equations.

$$h(t,X) = h_0(t)\, e^{\beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p} = h(t,X) = h_0(t)\, e^{\sum_{i=1}^{p} \beta_i x_i} \qquad \textbf{Equation 5-18}$$

where:

$h(t,X)$ is the hazard rate at time t;

$h_0(t)$ is the baseline hazard at time t;

p is the number of covariates;

β_j is the value of the ith regression coefficient;

x_i is the value of the ith covariate.

The hazard function h(t,X) is a measure of the potential for the event (failure, death) to occur at a particular time t, given that the event has not occurred yet (Kleinbaum and Klein, 2005, Cox and Oakes, 1984). It is also known as the failure rate, hazard rate, or force of mortality, etc.

The Cox Proportional Hazard (PH) model is usually defined in terms of the hazard model (Equation 5.18) expressing the hazard at time *t* for an individual i with a given specification of a set of explanatory, independent variables (also called *covariates* or *predictors*) denoted by X. The Cox model formulae involves a baseline hazard function, $h_0(t)$, which is a function of time t, but does not involve the covariates X's (that are time-independent variables). In contrast, the exponential expression involves the X's but does not involve t. This is indeed one of the advantages of the model.

The goal of the so called PH assumption is to evaluate if the hazard rate is constant over time, or equivalently, if the hazard for one individual is proportional to the hazard for any other individual, where the proportionality is constant over time (Kleinbaum and Klein, 2005; Cox and Oakes, 1984).

It is possible, however, to consider covariates that consider the time t. If those time-dependent covariates exist in the Cox regression model, the model will no longer satisfy the PH assumption and an *extended* Cox Model should be adopted. This extended model will consider simultaneously time-independent, X_i, (with coefficients β_0, $\beta_1,\ldots,$ β_{p1}), and time-dependent covariates, X_i (t), (with coefficients δ_0, $\delta_1,\ldots,$ δ_{p2}) as follows:.

$$h(t,X) = h_0(t)e^{\beta_0+\beta_1X_1+\ldots+\beta_{p1}Xp_{1+}\delta_0+\delta_1X_1(t)+\ldots+\delta_{p2}X_{p2}(t)} = h(t,X) = h_0(t)e^{\sum_{i=1}^{p1}\beta_i x_i + \sum_{i=1}^{p2}\delta_i x_i(t)}$$

Equation 5-19

Therefore, we have analysed the regression coefficients from a Cox extended model to estimate the impact of the covariates on the survival of organizations in our simulation. The cohort of firms[74] that has been considered was produced by the most fitted hypothesis in this simulation: we have gathered the results of 545 firms produced by CASOS simulation. We have used SPSS (Statistical Package for Social Sciences), version 15.0, (SPSS for Windows, 2007) to obtain the results that are presented in the subsequent analysis. We analysed the following covariates:

- *Size;*
- *DensCont*: contemporaneous density (number of firms at the time of a firms' death);
- *DensFound*: density at founding (number of firms at the time of a firms' birth).

Table 5-12 presents a summary of the data before processing the survival analysis. 360 cases were censored corresponding to the organizations that were still alive (state=1) at the end of the simulation.

		Number of cases	Percent
Cases available in analysis	Event	185	33.8%
	Censored	360	66.1%
	Total	544	99.8%
Total		545	100.0%

Table 5-12 - Case Processing Summary

[74] A cohort of firms is a group of firms who share a common characteristic or experience within a defined time period. In our study, firms within the cohort have in common the characteristic that they were all produced by the most fitted hypothesis in our simulation.

To proceed with our analysis, we test the Proportional Hazard assumption by calculating the correlations between the *Schoenfeld partial residuals* (obtained for each covariate) and the ranked order of survival time (Age) for not-censored observations[75]. If the correlations obtained with this process are significantly different from zero, it will mean that covariates are time-dependent and an *extended* Cox Model must be computed.

The partial residuals are quite symmetrically distributed, although we have confirmed that they were not normally distributed. Consequently, we have computed the *Spearman* correlations (more adequate for this kind of data) between the rank of *age* (the covariate that indicates the time of survival) and each covariate[76]. The first line of Table 5-13 shows the values of the Spearman correlation between the *Shoenfeld partial residuals* and the ranked order of survival time for non-censored observations.

		Rank of *Age*	Partial residual for *Size*	Partial residual for *DensCont*	Partial residual for *DensFound*
Rank of *Age*	Correlation Coefficient	1.000	-.650(**)	.106	-.045
	Sig. (2-tailed)	.	.000	.152	.546
Partial residual for *Size*	Correlation Coefficient	-.650(**)	1.000	-.148(*)	.022
	Sig. (2-tailed)	.000	.	.045	.766
Partial residual for *DensCont*	Correlation Coefficient	.106	-.148(*)	1.000	-.447(**)
	Sig. (2-tailed)	.152	.045	.	.000
Partial residual for *DensFound*	Correlation Coefficient	-.045	.022	-.447(**)	1.000
	Sig. (2-tailed)	.546	.766	.000	.
	N	184	184	184	184

Table 5-13 - Spearman correlations between the Shoenfeld partial residuals and the ranked order of survival time for non-censored observations

[75] The Schoenfeld residuals are computed for each covariate and for every case that are not censored. The value of the Schoenfeld residuals is the covariate value X_{ik} (where k indicates the covariate and i indicates the case) minus the expected value of the covariate for the risk set at t. This expected value is in fact a weighted average of the covariate, weighted by each individual's likelihood of dying at time t. Considering a set with J covariates, the partial residual for the case i, (PR_i), may be written as:

$$PR_i = x_{ik} - \sum_{i=1}^{J} x_{kj} p_j$$

[76] Instead of computing these correlations, we may analyse the plots of Schoenfeld partial residuals against time (rank of age) for each covariate to evaluate the Proportional Hazard assumption. In the case where there is a linear trend between Schoenfeld residuals and time, it indicates that the PH assumption is violated for that covariate.

Notes: ** Correlation is significant at the 0.01 level (2-tailed); * Correlation is significant at the 0.05 level (2-tailed).

The p-values for the correlation test are the p-values for the PH test in which the null hypothesis is that the PH assumption is not violated, as follows.

H_0: the PH assumption *is not* violated;
H_1: the PH assumption *is* violated.

The correlations are strong for *size*, since the p-value is lower than 1% (as indicated in the line called *Sig(2-tailed)*). It indicates that the PH assumption is violated for this covariates, since there is a significant association between the *Age* and the covariate. The covariate depends on the time and the Proportional Hazard Assumption is not supported for that covariate. An Extended Cox Model must be used and time-dependent variables will be created. We have considered two choices among the most usual forms for defining time-dependent covariates: (Kleinbaun and Klein, 2005)

- the product between the time and the covariate, *TimexCovariate;*
- the product between the natural logarithm of time and the covariate, ln(*Time*)x*Covariate.*

We tested both forms and concluded that there are no differences in the conclusions drawn from the results[77]. Therefore the time-dependent covariate corresponding to *size* is the product between the time the covariate: *Size (t) = Size x T*

Table 5-14 shows the omnibus tests of model coefficients (Kleinbaum and Klein, 2005; Cox and Oakes, 1984). This test allows for an overall evaluation of the model:

-2 Log Likelihood	Overall (score)		
	Chi-square	df	p-value.
1734,394	30.813	3	.000

Table 5-14 - Omnibus Tests of Model Coefficients in Cox Regression

[77] The simplest construct to build a time-dependent covariate is just the product between the time, T, and the covariate (Kleinbaum and Klein, 2005). We have used this construct in further time-dependent covariates. Comparisons between the results of the forms of constructs, made especially for the problem of Chapter 6 are presented in Appendix V

The likelihood ratio, the same as -2Log(Likelihood), and chi-square statistics given in Table 5-14 are asymptotically equivalent tests of the omnibus null hypothesis that all the coefficients β's are zero. In this case the null hypothesis is strongly rejected, (p-value<1%) indicating that there are covariates with a significant impact on the survival of organizations.

Table 5-15 shows the estimated coefficients of the covariates, $\hat{\beta}$, and the corresponding standard errors, represented as SE($\hat{\beta}$). The Wald statistic is computed through the quotient between the coefficient β and its standard error SE(β), and then taking the Normal (0,1) distribution for computing the corresponding value. The column *df* stands for the degrees of freedom involved in the test of hypothesis.

The exponentials Exp($\hat{\beta}$) measure the impact of the variables on the hazard of the firms. In fact, Exp($\hat{\beta}$) is the predicted change in the hazard for a unit increase in the predictor. The column named *p-value* shows the significance of the predictor concerning the test: **H_0: β=0 *versus* H_1: β≠0.**

Covariates	$\hat{\beta}$	SE($\hat{\beta}$)	Wald	df	p-value	Exp($\hat{\beta}$)
DensFound	-.089	.102	.772	1	.380	.915
DensCont	-.023	.025	.849	1	.357	.977
Size(t)	-.255	.045	32.637	1	.000	.775

Table 5-15 – Summary of covariates statistics of the extended Cox Regression Equation model[78]

According to these coefficients the extended Cox model, containing time dependent and time independent covariates, can be written as follows:

$$\hat{h}(t,X)=\hat{h}_0(t)e^{-0.023DensCont-0.089DensFound-0.255Size(t)}$$ **Equation 5-20**

[78] The Wald statistic presented in this table (and in others, later on) is calculated for each covariate in the model in order to measure the significance of the covariate on hazard rate. The Wald statistic for the covariate X_j is defined by $[\hat{\beta}_j/SE(\hat{\beta})]^2$, in which $\hat{\beta}_j$ is the coefficient estimate of covariate X_j and SE($\hat{\beta}_j$) stands for its standard error. Then, it is assumed that this quantity follows approximately a standard Normal distribution. For categorical covariates, a single significance is computed for the Wald statistic, taking into account the overall effect of all modalities of the covariate.

We have compared the sign of the coefficients we habe obtained with those obtained in other work (Mata and Portugal, 1994; Mata *et al.* 1995; Carroll and Hannan, 1989; Carroll and Hannan, 1992). The second column in Table 5-16 shows the predicted signs (obtained from literature; the same signs have been predicted in all studies) of the association between firm survival and the covariates.

Covariate	Predicted Sign	Obtained Sign
Density at Founding (*DensFound*)	+	-
Contemporaneous Density (*DensCont*)	-	-
Size	-	-

Table 5-16 – Predicted signs of coefficients obtained in literature and signs of the coefficients of the covariates obtained in the Cox regression extended model

Only *Size* has a significant impact on the hazard rate, which can be seen from the p-values. Carroll and Hannan (1989) explain that the density at founding has a positive impact on the mortality of organizations, and consequently on the hazard rate $\hat{h}$ **(t,X)**. In this case, the estimate of coefficient β for density at founding is negative. Concerning the other covariates, their impact is the same as previewed in literature, although the effect of contemporaneous density is not significant.

So, we can conclude that contemporaneous density (the number of firms existing at the moment of the death) has a negative, though not significant, effect on the mortality of organizations. As expected, the effect of the size on firm's mortality is also negative (and significant, in this case), confirming what was said before and helping to validate the simulation model.

Taking the column Exp($\hat{\beta}$) as the predicted change in the hazard for a unit increase in the predictor, we can state that, for instance, the value of Exp($\hat{\beta}$) for *Size*, whose value is 0,775, means that an increase of one unit on size decreases the failure hazard in 0,775 times, that is 22.5%.

5.6.3. Discussion: interpretation of simulation results and questions of validation

The performance of the system is rather consistent when compared to reality. We have observed similar tendencies between simulated and real data that represent the evolution of the number of industries in this region[79]. This provides an argument for model adequacy.

Furthermore, the signs of the covariates obtained with the survival analysis also confirm the capability of the model to reproduce the reality: the effect of the *size* on firm's mortality is negative and statistically significant, confirming what was said in literature. The sign of the covariate contemporaneous density *(DensCont)* is also negative as in literature, but with no significance.

In order to estimate the parameters of the density intervals, we have analysed the outputs of several runs of the simulation and retained some of the hypotheses that constitute the best solutions of the calibration process. The fitness that was used to determine the optimized solution is a combination of Manhathan and Growth distances (comparing simulated with real birth and death rates), as presented in Table 5-11.

These solutions of CASOS cannot be confirmed because there is no available data for this kind of analysis covering Portuguese industry. So, a question arises whether this set of parameters is unique and stable or, alternatively, whether there is more than one solution. Having analysed our results, we can affirm that the solution is not unique since there is some variation in the *combined* fitness for different solutions, as showed in Table 5-11. We saw that the fittest solutions are contained in the following integer intervals for the survival and founding densities, age and size[80]:

[79] This similarity was confirmed for *birth* rates and for the *total number of firms* in some of the regions. We could not confirm the similarity between the simulated and real *death* rates.

[80] We recall that DS_l and DS_u are respectively the lower and upper bounds of the density survival interval and DB_l and DB_u are respectively the lower and upper bounds of the founding (or birth) interval. A_l and S_l are the size, and age lower thresholds.

$DS_l=\{3:5\}$; $DS_u=\{10:28\}$; $DB_l=\{2:4\}$; $DB_u=\{14:25\}$; $S_l=\{1:4\}$; $A_l=\{1\}$ **Equation 5-21**

These solutions define the hypothetical limits of the survival and founding intervals, age and size for this industrial region concerning the seven-year period between 1997 and 2003[81]. New firms are almost always accepted (because density birth interval is wide), but only some of them will survive: infant firms are most likely to fail almost when they born (the Age limit is 1) and firms that attain the age of 1 will prevail. The size limit is more tolerant: the lower limit S_l stands between 1 and 4 meaning that firms whose size is greater than 4 are more likely to survive in this region. The density survival (DS) levels lay between 3 and 28. These values constitute a legitimation limit established for the region.

Analysing the real situation in what concerns the density (number of firms per square kilometrex10) in the recent past, we can observe that the real DS_l limit is a little higher, but the DS_u is within the interval (see Figure 5-13).

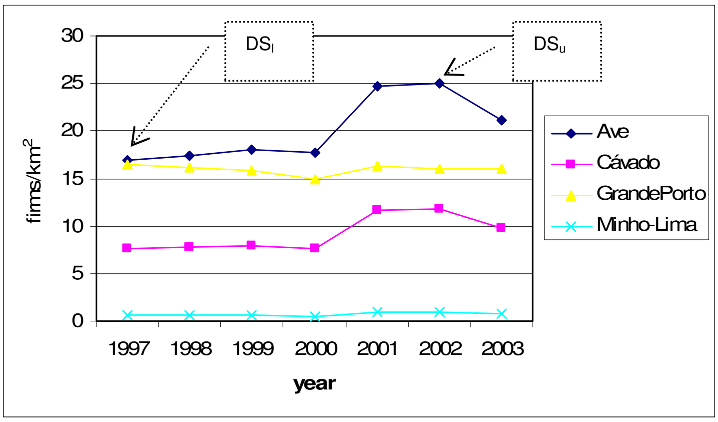


Figure 5-13 – Real evolution of density (firmsx10/Km^2) in the four regions of the study (DS_l and DS_u limits are showed up for the region of AVE)

All the regions of the study face a decrease in the organizational density, after a peak in 2002 and the value that corresponds to the top of the legitimation level for the region of Ave (between 2001 and 2002) is equal to 25.

[81] All values of the density interval limits (DS and DB) were multiplied by ten.

5.6.4. Conclusions and progress

Our aim with CASOS is to estimate the parameters of the density intervals that characterize the niche dimension for a particular industry. The outputs of several runs of the simulation were analysed and the hypotheses that constitute the best solutions of the calibration process were retained.

An exact final solution has not been found, since some slight variations exist in the parameters as we saw in Equation 5.23. The complexity of the interactions among cellular automata generates differences in the final fitness and it is not possible to prove that a particular solution is always associated with the same fitness. We faced another problem: the simulated death rates that intervene in the Fitness function of the calibration process differ much from the real ones.

Nevertheless, we obtained a range of solutions that are corresponding to reality: the solution of the DS_u (density survival upper limit) that establishes the legitimation limit is within the real interval that was observed in last years. This limit has a particular interpretation in the light of organizational ecology theories presented in Chapter 2. According to the theory, after reaching a certain level of growth, the maximum level of legitimation is reached, and competition between firms starts to increase fast. Consequently, the growth rate decreases fast to zero and the density falls down to very low values. We think that this limit has been reached in the region of Ave, and therefore the density will decrease fast.

For public and managerial policies this conclusion can help in the definition of political measures to stimulate the birth of new firms in the region, and to avoid others from disappearing.

Some model improvements can be made in the future. For example, an alternative for the calculation of the fitness of the hypotheses (to avoid the great divergence between simulated and real death rates) can be achieved by minimizing the maximum of error between rates. Therefore, instead of minimizing the half sum of the distances, as we saw in Equation 5-6, we could minimize the maximum of the distances:

$$\varphi(h_i) = MinMax\{d_b, d_d\}$$ **Equation 5-22**

This is actually a common method of Minimax to prevent aberrant values (namely in what concerns the death rates produced by the simulation) and can be used as an alternative to the computation of the distances that was used above.

In the next chapter we present the second study, NetOrg, a more detailed Multi-Agent framework to analyse the dynamics of organizational survival in collaboration networks.

6. A Multi-Agent System for collaboration networks

Introduction

In this chapter, we propose a Multi-Agent framework to analyse the dynamics of organizational survival in cooperation networks. This approach uses microeconomic formalization and Multi-Agent systems.

One of our aims is to study the effect of different collaboration[82] strategies on the survival of firms. As we saw in the previous chapters, firms may share some resources and develop a cooperative behaviour in what concerns, for instance, research and development activities that relates with innovation. Innovation capabilities and

[82] The term *collaboration* is used to describe the main motivation for the association between firms. Sometimes *cooperation* and *collaboration* are used indistinctively, both indicating a certain type of assotiation between firms, although there is a clear difference between the two concepts, as it was introduced above in Chapter 3. The distinction between the two concepts is explored again later in Section 6.9.

technological sophistication are crucial in strategic alliances, so the connection between innovation and networking is of great importance[83].

In this chapter, we assume that firms may cooperate *horizontally* (in the same market) or *vertically* (with firms from other markets that may belong to the same supply chain). Cooperation decisions are based on cognitive and economic variables. A variant of the density dependence model has also been exploited here. To validate our model, we have used data obtained in previous studies presented in Chapter 4, and tested some hypotheses assotiated to realistic assumptions.

We have observed that firms and networks proliferate in the regions with lower marginal costs, but new networks keep appearing and disappearing in regions with higher marginal costs. Simulatenously, we concluded that the diameter of networks decreases along time and transitiviy (the clustering coefficient of networks) tends to increase strongly in the inital life of neworks and decreases slowly in the end.

In the following sections (6.1 to 6.8) we introduce the modelling approach, that is, the model components and their function as well as technical aspects of the simulation algorithm, namely the configuration of model parameters and implementation. We introduce the issues concerning the development and validation of the Multi-Agent System (MAS). The results and concluding remarks are given in Sections 6.9 and 6.10.

6.1. The Modelling Approach

According to evolutionary economics, firms innovate in order to increase their chances of survival. In fact, and as Schumpeter (1996 [1942]) emphasizes, innovation is a powerful vehicle for new firms to successfully enter the market and undermine the established firms. As well, established organizations need innovating to maintain their competitive position in the face of new and emerging or 'disruptive' technologies. Also, Baumol (2002) states that under capitalism, innovative activity becomes mandatory, a life-and-death matter for the firm and innovation has replaced price as the name of the game in a number of important industries.

[83] As introduced in Chapter 2, networks of knowledge, innovation and technology (KIT) are examples of networks that allow firms to have access to technological knowledge and innovation.

Some authors have related the process of innovation with the creation of networks of firms, using simulated environments: Carayol and Roux (2003) study innovation as a collective and interactive process that generates the formation of networks of organizations [84] . Wersching (2005) uses an Agent-Based model to study the technological development, the economic performance of firms and the evolution of agglomerations in differentiated industry[85].

In this work, we adopt the approach of Agent-Based computational economics (Tesfatsion, 2006), by assuming that the computational study of the economic processes is modeled as dynamic systems of interacting agents. Alternatively, these problems could be partially solved by optimization, as it is currently done in mainstream economics.

Using a MAS framework, where firms (the agents) may cooperate, we propose to analyse the dynamics of collective innovation. This framework is appropriate because we can simulate agents that can be configured to be autonomous, enabling to capture the dynamics (and the survival) in network formation through cooperation.

Therefore, we have developed *NetOrg*, the Multi-Agent framework where the agents interact and constitute organizational networks for the purposes of innovation. This framework includes a microeconomic modelling, several strategies for cooperation, a cognitive model and a process of decision making.

6.2. Model components

6.2.1. The environment

The environment includes a SxS grid where the firms are located (each firm in this grid has two coordinates). We have defined that S, the dimension of the square grid, is equal

[84] However, the approach from Carayol and Roux (2003) does not include many aspects as cognitive modelling and learning as we are going to implement in this essay.
[85] We could mention some other works, as Zhang (2003), Gilbert et al. (2001), Cowan et al. (2004), Andras et al. (2006), Cortés and Sheremetov (2002), just to get a good overview of the application of those models. For studies on modelling the network structure, see Purchase and Olaru (2003).

to ten times the number of firms at the beginning of the simulation. Therefore, there will be enough room for the creation of new firms in the organizational environment. As we will see later, the number of initial firms is endogenous in the Multi-Agent System, being dependent on other parameters. The grid is divided in two regions: one with higher marginal costs, *Region 1*, and another with lower marginal costs, *Region 2* (see

Figure 6-1). Firms get born and die according to the levels of density and other parameters that are defined further in this section. Firms can also migrate between the two regions.

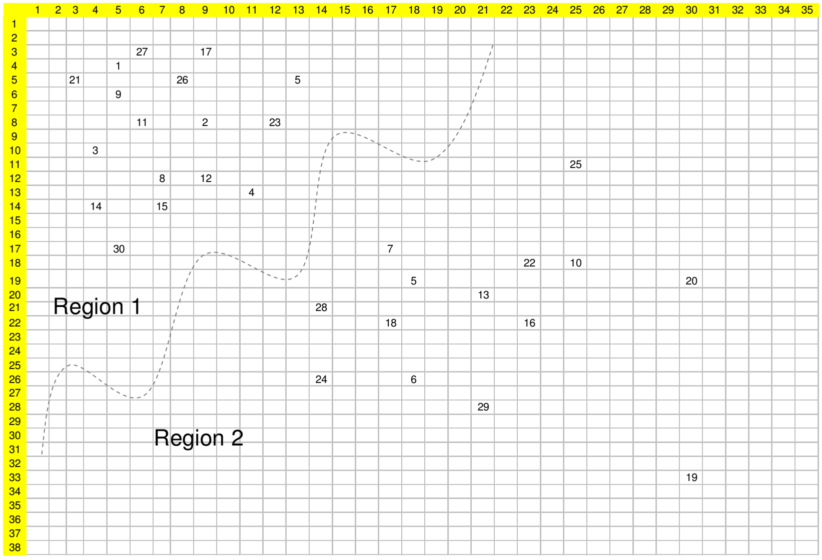


Figure 6-1 – Space of *NetOrg* represented by a SxS grid

Note: the dotted line divides two regions: one with higher marginal costs, Region 1, and another with lower marginal costs, Region 2

6.2.2. The agents

We considered three markets (X, Y_1 and Y_2). X is the final good market and Y_1 and Y_2 are the intermediate good markets. The market X contains **n** firms, while markets Y_1 and Y_2 contain, respectively, $\mathbf{m_1}$ and $\mathbf{m_2}$ firms. The total number of firms is $\mathbf{N=n+m_1+m_2}$. Car makers, carburettors and clutches industries are examples of these three types of markets (as in Swaminathan, 2002).

6.3. Microeconomic foundations

6.3.1. Networks, knowledge creation and diffusion

According to Watts and Strogatz (1998), a network of firms may be represented as a graph containing vertices (nodes) and edges (links, arcs, or connections)[86]. The nodes represent the firms in the network, while links represent the connections for the purposes of knowledge diffusion.

For every market (X, Y_1, Y_2), we consider a different kind of knowledge (or stock of capital) represented by k^t_i, the stock of knowledge owned by firm i (i ∈ X, Y_1 or Y_2) at time t that is necessary to produce its product. Links between the firms allow for the diffusion of knowledge (the technological know-how) among the firms in the network. The decision of establishing a link and the flow of the knowledge inside the network is weighted by the distance between the firms, as we will see later in this section.

In this work, the nodes are not always connected (networks need not to be fully connected), that is, some part of the firms are linked to other firms. Figure 6-2 illustrates an example of a graph where four firms of different markets interact in a network:

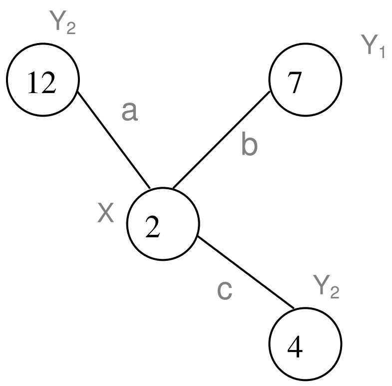


Figure 6-2 - Example of a graph representing a network of firms in markets X, Y_1, and Y_2.

Note: nodes 12 and 4 represent sellers in market Y_2, while node 7 represent a seller of market Y_1. Node 2 represents a buyer in market X. In further representations the identification of the markets and links will be deleted for illustration simplicity.

[86] In our model, although links have all the same weight, the distance between nodes is implicitly accounted for. In addition, our graphs are undirected (meaning that there is no implicit direction in the links).

We may relate the basic network of Figure 6-2 to a spatial representation in the grid as follows:

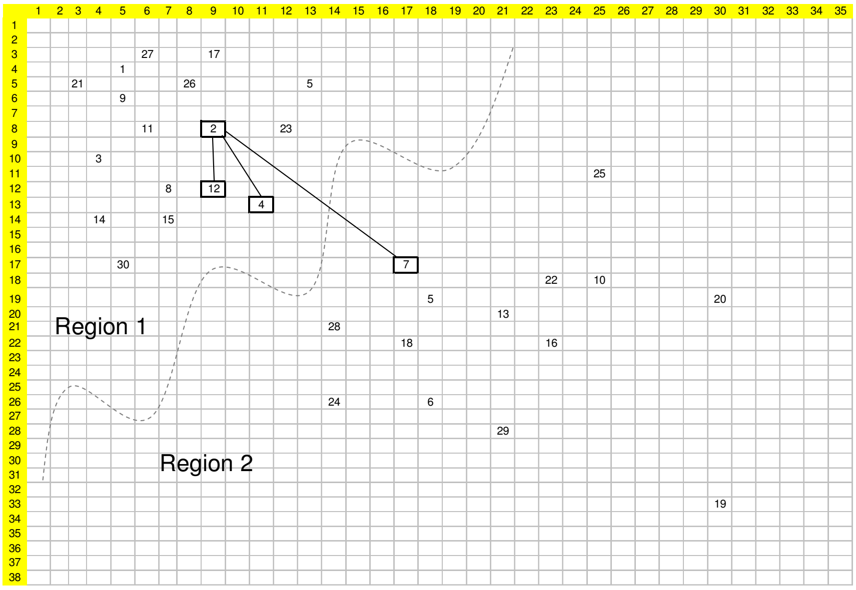


Figure 6-3 – Representation of a network in *NetOrg*'s grid space

The links can be represented in a (symmetric) square matrix (or *adjacency matrix*) in which the intersection between two nodes is either zero, if there is no link between them, or a positive number representing the identification of the link, otherwise. For simplicity, we will not use the identification of the link, which will be just replaced by the value 1.

	2	4	7	12
2	0	c	b	a
4	c	0	0	0
7	b	0	0	0
12	a	0	0	0

Table 6-1 – Adjacency matrix of the links in the graph of Figure 6-3

Knowledge creation and diffusion is the main goal of the network. Once a link between two firms is created, there is an immediate knowledge transfer among the firms in the network according to some rules (we will exploit the decision associated with the creation of links later in Section 6.3.2). Therefore, for every firm i, at instant t, the accumulation of knowledge k at time t is given by:

$$k_i^t = k_i^{t-1} + \Delta k_i^t \ , \qquad \textbf{Equation 6-1}$$

in which Δk^t_i represents the variation of the stock of knowledge of firm i ($i \in X$, Y_1 or Y_2) at time t.

Additionally, we have:

$$\Delta k_i^t = w_i^t + \sum_{j \in N \backslash i} \delta^{d^t(i,j)} w_j^t \ , \qquad \textbf{Equation 6-2}$$

in which w_i^t denotes the innovation of firm i ($i \in X$, Y_1 or Y_2) at time t and δ is the transferability factor (or spillover[87]), i.e., the parameter that measures the share of new knowledge which is effectively transmitted through the network. We denote by $d^t(i,j)$ the total distance between firms i and j at time t.

We assume that w^t_i is a function of the knowledge previously owned by firm i with a certain probability, p, specifically:

$$w_i^t = f(k_i^{t-1}, p) = k_i^{t-1} p \qquad \textbf{Equation 6-3}$$

in which p is particular value of a variable P having a Normal distribution with the following characteristics:

$$P \sim N(\lambda, 0.1\lambda) \ . \qquad \textbf{Equation 6-4}$$

Additionally, we will assume that λ differs according to each product or market.

[87] According to Scitovsky (1954), spillovers (or technological externalities) deal with the effects of non-market interactions, being realized through processes that affect the production (or profit) function of a firm. Spillovers may respect to the diffusion of learning across firms, which can take place through interfirm mobility of employees or cooperation.

We consider, as in Carayol (2003), that there is no knowledge transmission (or diffusion) between firms that do not share a network connection. In other words, knowledge is only transferred through links that are established. Therefore:

- if firms i and j do not share a network connection then: $\delta^{d^t(i,j)} = 0$.
- if firms i and j share a network connection then: $\delta^{d^t(i,j)} > 0$.

Additionally, we will assume that each firm in a particular market owns a stock of knowledge of its type, but may also have some skills on other types of knowledge. For instance, a final good producer (e.g. car makers) detains a specific stock of knowledge in its industry but may also hold some knowledge about intermediate good industries (e.g. carburettors or clutches suppliers). For each type of market, this knowledge is associated either with the product itself or with the production process.

The networks discussed above may be generically called collaboration networks, as there is, in some cases, a common goal among the firms in the network concerning knowledge creation and dissemination. However, as we will see later, this common goal may not be explicit in the network. In the next section we introduce collaboration networks and the process of network creation.

6.3.2. Cooperation/collaboration networks

In our study, collaboration networks emerge from the need of creating and diffusing knowledge concerning R&D (Research and Development) activities. R&D can be defined as any project to answer to scientific or technological uncertainty meant to accomplish an advance in science or technology. Advances include new or improved products, processes or services. In general, R&D activities are performed by specialized units or centres belonging to firms, state agencies or universities.

Collaboration networks may proceed by adopting a vertical or a horizontal connection. As we have said before, we assume that horizontal links are connections between firms within the same market, while vertical links are connections between firms from different markets. All vertical and horizontal links are called *collaboration links* or *cooperation links*, as they aim at relating firms in order to create and diffuse technological know-how. Besides, there are other types of collaborations, such as *cooperation in production*, which we will not study in this work, as they are seldom observed. Furthermore, organizations establish *production links* in the normal production process of the supply chain concerning buyer/seller relationships or seller/seller relationships.

As seen above, we consider three types of markets where X is the final good market and Y_1 and Y_2 are the intermediate good markets. In the final good market, there are consumers and firms (or final good producers), which, at the same time, are buyers at the intermediate good market. We can therefore consider the existence of three layers in this economic environment:

Intermediate goods producers (or sellers)

Final good producers (or buyers)

Consumers

Figure 6-4 shows a schematic example of a network, including examples of both types of connections and indicating the type of firms (buyers, sellers).

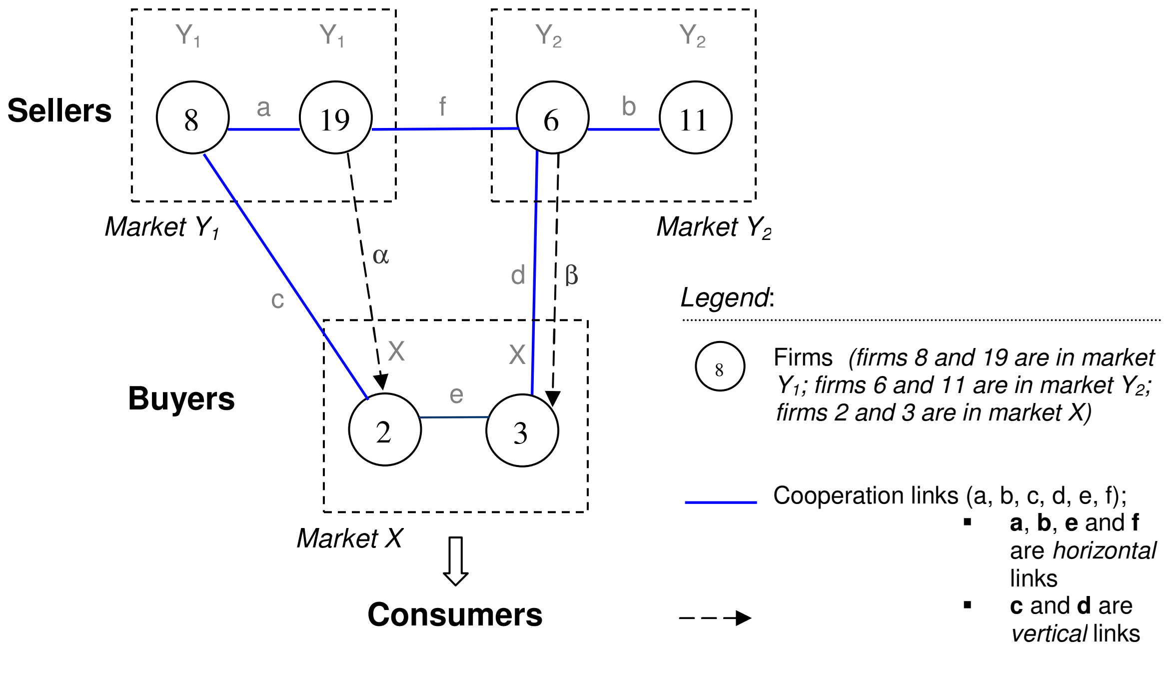


Figure 6-4 – Schematic representation of a collaboration network with both horizontal and vertical links

The evolution of networks is a matter of interest in our work, as one of our aims is to analyse the effect of cooperation/collaboration on the survival of firms. Therefore, we have followed the evolutionary paths of the networks and registered the entry and exit of firms in the network. In the following figures, networks are illustrated with *cooperation links* only. We just represent the *cooperation links* as they are the matter of concern in this work.

Figure 6-5 illustrates the evolution of a particular network. At the beginning (we call it Step 1 or generation 1), it contains three firms (2, 4 and 7) from three different markets. The network included two vertical links connecting these three firms. In generation 2 one new firm from market Y_2 (firm 12) joins the network. In the third generation, firm 4 left the network and other firm (firm 21) entered. The evolutionary path would continue until the disintegration of the network.

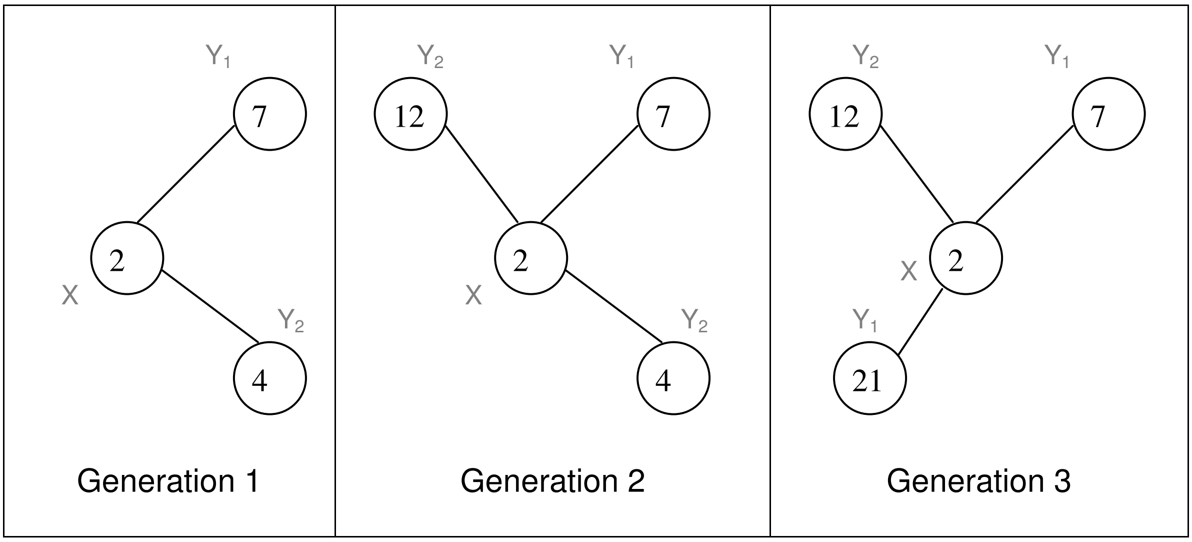


Figure 6-5 - Example of some generations of a network evolutionary process

6.3.3. Technological and geographical space

One of the most important aspects related with collaboration networks for innovation purposes is the distance between firms. Although the geographical distance is important for knowledge transmission (e.g. Jaffe, 1989; Jaffe *et al.* 1993; Audretsch and Feldman, 1996), the non-geographical distance between firms may also be relevant. Some authors (e.g. D'Agata and Santangelo, 2004) call it cognitive or technological distance and it plays a major role in the effectiveness of knowledge transmission that can be generated in inter-firm networks. This technological distance represents the difference of the technological know-how between two organizations. Sometimes, this difference cannot be calculated explicitly, and so economists use proxy indicators to obtain a value close to the technological distance. One of these proxies is the stock of knowledge[88].

Therefore, we considered two types of distances that were combined into one weighted distance:

(i) $\mathbf{d^{geo}(i,j)}$ represents the geographical distance between firms i and j, which is measured by the Euclidean distance;

(ii) $\mathbf{d^{tech,\,t}(i,j)}$ represents the technological (or cognitive) distance between firms i and j at time t, which is measured by the difference of firms' stock of knowledge.

More precisely, the geographical distance is computed as a normalized Euclidean distance where the distance between the coordinates of two firms (i and j) is divided by the maximum distance between any two firms:

$$d^{geo}(i,j) = \frac{\sqrt{(i_x - j_x)^2 + (i_y - j_y)^2}}{Max(d^{geo}(i,j))}, \forall i \neq j \;. \qquad \textbf{Equation 6-5}$$

where (i_x, i_y) are the coordinates of firm i in the grid and (j_x, j_y) are the corresponding coordinates of firm j. Max $(d^{geo}(i, j))$ is the maximum distance between any two firms of the population. Therefore, due to the definition used, $d^{geo}(i, j)$ has the following property in our simulation:

$$0 < d^{geo}(i,j) \leq 1, \;\; \forall i \neq j \qquad \textbf{Equation 6-6}$$

[88] The concept of *stock of knowledge* is related to the stock of capital, that is, to the investment in skills and education or intellectual property and also investments in equipments, machinery, buildings, etc. For that reason, in this research, stock of knowledge and stock of capital are used as synonyms.

d^{geo} is never null (otherwise two firms would have the same coordinates and occupy the same space in the grid) . The technological distance, d^{tec} is computed as follows:

$$d^{tech} = \begin{cases} \dfrac{\left|k_i^t - k_j^t\right|}{Max(K_\bullet^t)}, & \text{if } \left|k_i^{t-1} - k_j^{t-1}\right| \in [a', b'], \forall i \neq j \\ 0 & , \text{if } \left|k_i^{t-1} - k_j^{t-1}\right| \notin [a', b'], \forall i \neq j \end{cases}$$ **Equation 6-7**

where k_i^t is the stock of knowledge owned by firm i at time t and $Max(k_\bullet^t)$ is the maximum stock of knowledge owned by any firm. Additionally, we assume that firms do cooperate if the difference between their stocks of knowledge is neither too large nor too small. In fact, and according to D'Ágata and Santangelo (2004), a small cognitive distance allows greater comprehensibility, but yields redundant, novel knowledge. Conversely, a large cognitive distance allows limited comprehensibility, although yielding non-redundant, novel knowledge. Therefore, a certain degree of cognitive distance is needed since it ensures that firms can connect their cognitive frameworks and being innovative as well as they can easily communicate between each others.

Actually, two limits were defined to deal with this restriction, a' and b', which are defined exogenously in the simulation. Equation (6.7) shows that d^{tech} is a normalized distance, satisfying the following property:

$$0 \leq d^{tech}(i,j) \leq 1, \ \forall i \neq j.$$ **Equation 6-8**

It is important to note that the technological distance between two firms may be null. The final weighted distance is obtained by the equation:

$$d^t(i,j) = [W_{tech}\, d^{tech,\, t}(i,j)] + [W_{geo}\, d^{geo}(i,j)].$$ **Equation 6-9**

Without loss of generalisation, we have assumed here that $W_{tech} = 0.75$ and $W_{geo} = 0.25$, which gives more importance to the technological distance than to the geographical one.

Example: Consider two firms, 1 and 2, which have stock of knowledge of 21 and 23, respectively, while $Max(k_\bullet^t) = 5$. Therefore, the technological distance is 0.4. If the geographical distance between them is 0.1, then the total distance between firms 1 and 2 is 0.175:

$$d^t(1,2) = [0.75 * 0.2] + [0.25 * 0.1] = 0.175. \qquad \textbf{Equation 6-10}$$

6.3.4. Market demand

We considered a final good industry consisting of **n** firms, each producing a variant of the product. Consumers are assumed to have *love-for-variety* preferences (Dixit and Stiglitz, 1977), where the representative consumer has the following utility function:

$$u^t(x^t_1, \ldots, x^t_n) = \left[\sum_{i=1}^{n} \left(A^t_i \, x^t_i \right)^b \right]^{1/b} \qquad \textbf{Equation 6-11}$$

where A^t_i denotes the current attractiveness of final product variant i and $x^t_1, \ldots, x^t_n$ the consumption of each variety of final good. The parameter b stands for the inverse of the intensity of *love for variety* over the differentiated product. When b is close to one, varieties are close to perfect substitutes; when b decreases, the desire to spread consumption over all varieties increases.

The standard *love-for-variety* approach assumes equal attractiveness of the variants. In our approach, we will assume that the attractiveness A^t_i depends on the accumulation of knowledge, that is, consumers are more attracted to varieties of the final products that are more innovative:

$$A^t_i = \Delta k^t_i \qquad \textbf{Equation 6-12}$$

in which Δk^t_i represents the variation of the stock of knowledge of firm i ($i \in$ X, Y_1 or Y_2) at time t, as shown before in Equation 6-1. The utility function is maximized subject to the budget constraint (Varian, 1992):

$$\sum_{i=1}^{n} p^t_i \, x^t_i \leq R^t \qquad \textbf{Equation 6-13}$$

where R^t is the overall amount of money allocated by consumers at time t to purchase goods produced by this industry and p^t_i is the unit price that consumers pay for variety i at time t. Straightforward calculations yield the following inverse demand function (or price) for market i:

$$p_i^t = \frac{(A_i^t)^b}{(x_i^t)^{1-b}} \cdot \frac{R^t}{\sum_{r=1}^{n} (A_r^t . x_r^t)^b} . \qquad \textbf{Equation 6-14}$$

6.3.5. Cost structure and profit function

Final-good producers (or buyers) transform intermediate goods, Y_1 and Y_2, through a fixed coefficient technology[89].

Therefore, the unit cost of production of each final-good producer (or buyer) is then given by:

$$c_i^t = c + t_1 + t_2 - \Delta k_i^t$$

$$\Leftrightarrow \quad c_i^t = = c + t_1 + t_2 - w_i^t - \sum_{j \in N \setminus i} \delta^{d^t(i,j)} w_j^t \qquad (i = 1, \ldots, n) \qquad \textbf{Equation 6-15}$$

where **c** accounts for stand-alone marginal costs[90] that are analogous for all buyers in the same region[91] and Δk_i^t is the variation of the stock of knowledge explained above.

Buyers' profit function is therefore:

$$\pi_i^t = (p_i^t - c_i^t).x_i^t - g(w_i^t) \qquad (i = 1, \ldots, n) \qquad \textbf{Equation 6-16}$$

where $g(w_i^t)$ is the cost needed to create knowledge. As it is typical in R&D literature (e.g. d' Aspremont and Jacquemin, 1988), we assumed that there are diminishing returns to knowledge expenditures, that is, $g'(\cdot) > 0$ and $g''(\cdot) > 0$. Without loss of generality, we will assume that $g(w_i^t) = 0.5(w_i^t)^2$.

Similarly, we assume that the inverse demand functions that intermediate-good producers (or suppliers) face are:

---

[89] A fixed coefficient technology implies that inputs Y_1 and Y_2 are combined into a constant ratio to one another. That is, each unit of X uses a fixed quantity of both intermediate goods Y_1 and Y_2, which are perfect complementary (Besanko and Braeutigam, 2005).

[90] Marginal costs take under consideration the production technology and the prices of the production factors, such as price of capital and price of labour.

[91] We have defined two regions to which we have associated marginal costs: region 1 has higher marginal costs than region 2.

$$t_1^t = \sum_{j=1}^{m_1} \frac{A_j^t}{y_j^t}$$ **Equation 6-17**

for market Y_1 and

$$t_2^t = \sum_{j=1}^{m_2} \frac{A_j^t}{y_j^t}$$ **Equation 6-18**

for market Y_2.

The sellers' unit production cost is given by:

$$c_i^t = c_1 - \Delta k_i^t = c_1 - w_i^t - \sum_{j \in N \setminus i} \delta^{d^t(i,j)} w_j^t \qquad (i = 1, \ldots, m_1)$$ **Equation 6-19**

$$c_i^t = c_2 - \Delta k_i^t = c_2 - w_i^t - \sum_{j \in N \setminus i} \delta^{d^t(i,j)} w_j^t, \qquad (i = 1, \ldots, m_2)$$ **Equation 6-20**

where $\mathbf{c_1}$, $\mathbf{c_2}$ accounts for stand-alone marginal costs, which are similar for each type of sellers (Y_1, Y_2) in the same region.

The sellers' profit function is then given by:

$$\pi_i^t = (t_1^t - c_i^t).y_i^t - g(w_i^t) \qquad (i = 1, \ldots, m_1)$$ **Equation 6-21**

$$\pi_i^t = (t_2^t - c_i^t).y_i^t - g(w_i^t), \qquad (i = 1, \ldots, m_2),$$ **Equation 6-22**

where $\mathbf{y_i^t}$ represents the quantity produced by firm i at instant t.

6.4. Decision making

6.4.1. Entry/Exit and Organizational Survival

The entry and exit in the market (birth and death processes) were defined by a variant of the density dependence model (Campos and Brazdil, 2005). According to this variant, which was already discussed in Chapter 5, the chances of a firm to get born or die depend on the number of firms that are located in its neighbourhood.

So, the firm's entry and exit will be modelled in the same manner that it was made in CASOS: if the number of living organizations in the neighbourhood of a specific firm belongs to the survival interval [DS_l ; DS_u], then the organization will have a higher probability to stay alive[92]. Otherwise it will tend to die by the effect of "overcrowding" or "solitude", depending on whether the number is greater than DSu or lower than DSl, respectively. The same idea applies to the process of founding.

Therefore, we have defined (in Equation 5.4), the following rules that determine the state (δ_{ij}^t), of a firm i in region j:

$$\delta_{ij}^t = \begin{cases} 1 \text{ if } (DS_l < f_i^t < DS_u) \wedge (S_i^t < S_l) \wedge (A_i^t < A_l) \wedge \delta_{ij}^{t-1} = 1 \\ 1 \text{ if } (DB_l < f_i^t < DB_u) \wedge \delta_{ij}^{t-1} = 0 \\ 0 \text{ otherwise} \end{cases} \qquad \textbf{Equation 5-3}$$

We have also considered the same birth and death (entry and exit) function for each of the firm attributes - density survival, size, and age – that was presented in the Figure 5-5, for which a probability of survival can be computed.

Concerning the death of a firm (δ_{ij}^t=0), some other conditions have been introduced: a firm also dies if in addition to the conditions presented in Equation 5.3, one of the following situations happens:

- Firm dies if it stays more than three consecutive periods in the situation of having a negative profit:

  $$(\pi_i^{t-1} < 0 \wedge \pi_i^{t-2} < 0 \wedge \pi_i^{t-3} < 0) \qquad \textbf{Equation 6-23}$$

- Firm dies if it is not recovered by the *Network support* function. Firms with small profit (more than three consecutive periods with negative profits) would normally fail, unless they have been recovered by their neighbours based on a function called *Network support*. This function will be explained later in Section 6.3.7.

The conditions for new firms to be born are the same of CASOS: if the number of living organizations in the neighbourhood of a specific firm belongs to the founding interval [DB_l ; DB_u], then a new organization will be born. Nevertheless, a new condition has

[92] As in Chapter 5, the parameters are defined by the same symbols: DS_l and DS_u are, respectively, the lower and upper bounds of the density survival interval and DB_l and DB_u are, respectively, the lower and upper bounds of the density founding interval.

been imposed for firms' birth to take place in a certain neighbourhood (defined later as a parameter of the simulation): the mean profit of the neighbouring firms must be positive.

A function named *location*[93] was implemented in order to determine the exact location (coordinates) of the new firm in our simulated space. The new firm will be associated with a certain product (X, Y_1 or Y_2), and will hold the stock of knowledge, k^t_i. In our simulation a random process defines all these parameters.

6.4.2. Creation and spread of innovation

As stated before, innovation at time t, defined by w^t_i depends on the knowledge accumulated by the firm (Equation 6-3). In line with empirical evidence, we assumed that all firms with positive profits would invest in R&D. That is, if $\pi_i^{t-1} > 0$, for a particular period time t, then the innovation w_i^t will not be null.

We have considered that in all generations of the evolution of networks, innovation is increased by a value, p, coming from a Normal distribution. This model is in line with the work of Carayol and Roux (2003), in which networking allows the flow of R&D innovation to spread within a network and, consequently, to increase the stock of knowledge of other firms that are connected in the same network.

6.4.3. Collaboration/Cooperation Networks

Cooperation is considered to be a bilateral process where a single firm can choose only one partner (another firm) to cooperate with. So, the decision to cooperate is determined by the two firms involved. In some cases, we can admit that a group of firms (more than two) can join together at the same time and form a cooperation network.

[93] The birth process is associated with the density in the region of a particular firm i. Therefore the new entrant firm is located close to firm i. Details about the implementation of the simulation algorithm, NetOrg, will explained in Sections 6.5 to 6.7

In the description of the strategies, refer that they are neither complementary nor alternatives. In addition, the decision of the strategy to use does not depend on the firms. The choice of the strategy is exogenous in the simulation. We use different strategies and compare their impact on economical indicators (such as profit and knowledge stock) and on the survival of the firms.

There are two phases concerning the cooperation process in which firms can link together to form a network: the *selection of the partner* and the *network formation*. In the following we give several alternatives of these phases as different possibilities that firms can pursue in order to initiate cooperation. After that, in Section 6.9, these phases are combined to form different cooperation *strategies*[94].

A. Selection of the partner:

A1. Preferential meeting process

The selection of the partner in the network can be done by the *preferential meeting process* (Carayol and Roux, 2003; Campos, *et al.* 2006). In the *preferential meeting process,* we have defined a *meeting affinity*, $m^t_{i,j}$, between two firms, in which firm i chooses the firm j that maximizes the value $m^t_{i,j} = [1-d^t(i,j)]$. In this expression, $d^t(i,j)$ represents the pooled (geographical and technological) distance between i and j. By this approach, we intend to capture the influence of both geographical and technological distance on the formation of networks, which is evidenced in empirical research (e.g. Cassiman and Veugelers, 2002).

A2. Average stock scenario

Since the profit of firm i at time t depends on its stock of knowledge (k^t_i) - which is improved by cooperation -, an alternative to the preferential meeting process is to select a partner that is already linked with other firms (the most linked one). After that, its expected growth is computed to see if the stock of knowledge is increasing. This

[94] The several alternatives considered in these two phases have been inspired in the literature (an overview of previous work is present later on). Strategies are neither complementary nor alternative. In addition, the strategy to adopt does not depend on the decisions of firms. We define different strategies independently and compare their impact on economical indicators (such as profit and knowledge stock) and on the survival of firms.

process is based on the perspective of rational choice from Simon (1955)[95] and is systematized as follows:

- The most linked firm is selected as a potential partner[96];
- Then, to decide about the formation of a network, we compute the increase on the stock of knowledge in the following way:

(i) Let j be the selected potential partner to form a network with firm i in this process; let k_j^{t-1} be the stock of knowledge of j at ime t-1 (last generation). Then, the rate of variation of the stock of capital for firm j between time t-1 and t-2 is given by:

$$\Delta v_j = \frac{k_j^{t-1}}{k_j^{t-2}} - 1$$ **Equation 6-24**

(ii) Two expected scenarios for the increase of the stock of knowledge can then be computed by firm j for the actual period (t and t-1): either the increase is greater than Δv_j (let's say, 10%), or it will be lower than Δv_j (again 10%). An estimate of the average scenario is made by pooling the two situations with a factor ϕ_1:

1) $k_j^{t-1}\ \Delta v_j * 1.1$ **Equation 6-25**

2) $k_j^{t-1}\ \Delta v_j * 0.9$ **Equation 6-26**

As a result, in the average scenario, the growth of the stock of knowledge at time t, defined by $E(k_j^t)$, is computed as the expected value for a discrete variable, as follows:

$$E(k_j^t) = \phi_1(k_j^{t-1}\ \Delta v_j \cdot 1.1) + (1-\phi_1)(k_j^{t-1}\ \Delta v_j \cdot 0.9)$$ **Equation 6-27**

[95] This strategy could be called *Reinforcement strategy* as individuals tend to adopt actions that yielded a high payoff in the past, and avoid actions that yielded a low payoff. This is the standard learning model in behavioural psychology that has gained the attention of economists. As in imitative models, payoffs describe choice behaviours but it is "*one's owns past payoffs that matter, not the payoffs of others*" (Young, 1998), as we have seen back in Chapter 3.

[96] If more than one *most linked firm* exist, then one is chosen at random among firms in that condition.

Therefore, the decision of cooperation between firms i and j is based on the comparison between the expected value, $E\,(k_j^t)$, and the growth rate of the stock of knowledge of firm i, Δv_i, as we given later in B.1.2.

B. Network formation

B1. Individual decision for integration in new/previous network

We considered several forms of decision for the integration of a firm in a network: *resource complementarity*, *expected growth of the stock of capital* and the *concentration process.*

B1.1. Stock ratio complementariness

Stock ratio complementariness is a simple process in which two firms collaborate if they are able to reciprocally compensate some lack of profitability in the stock of knowledge (measured by the ratio π_i^t / k_i^t) concerning some product. This process is inspired in the common sense idea of cooperation. For example, if firm i has higher profitability in the stock of knowledge than firm j, regarding product Y_1, but firm j has higher profitability in the stock of knowledge than firm j regarding product Y_2, then they can collaborate in exchanging knowledge concerning a new kind of product or process that may increase the quantities they both produce. This process can be defined by the following rule:

$$\text{if } (\pi_i^t / k_{i,Y_1}^t) > (\pi_j^t / k_{j,Y_1}^t) \text{ and } (\pi_i^t / k_{i,Y_2}^t) < (\pi_j^t / k_{j,Y_2}^t) \text{ then} \Rightarrow$$

$$\Rightarrow \text{firms i and j collaborate on } Y_1, Y_2\,. \qquad \textbf{Equation 6-28}$$

In this example, $k_{i,\mathbf{Y_1}}^t$ represents the stock of knowledge of firm i at time t concerning specifically product Y_1. Generically $k_{i,\mathbf{Y_s}}^t$ represents the stock of knowledge concerning product s (with s = X, Y_1, Y_2). In the following table we give two examples in which we observe the conditions that drive to the collaboration between two firms:

	Cooperating firms		Product		profit / stock of knowledge				Region	
#	Firm i	Firm j	Firm i	Firm j	$\pi^t_i / k^t_{i,Y1}$	$\pi^t_j / k^t_{j,Y1}$	$\pi^t_i / k^t_{i,Y2}$	$\pi^t_j / k^t_{i,Y2}$	Firm i	Firm j
1	1	24	1	3	25.73	40.56	44.83	35.62	2	1
2	2	24	1	3	16.31	22.32	28.05	19.56	2	1

Table 6-2– Examples (identified as #1 and #2) of two firms that collaborate in *stock ratio complementariness*

In this example, the first line represents a situation in which firms i = 1 and j = 24 (producing products 1 and 3, respectively), have a complementary ratio of Profit / Knowledge Stock. In fact, considering product Y_1, the ratio for firm i is smaller than the corresponding ratio for firm j ($\pi^t_i / k^t_{i,Y1}$ = 25.73 while for firm j is $\pi^t_j / k^t_{j,Y1}$ = 40.56). Meanwhile, considering product/process Y_2, the situation is opposite: the ratio for firm i is greater than the corresponding ratio for firm j: $\pi^t_i/k^t_{i,Y2}$ = 44,83 while for firm j the corresponding ratio is $\pi^t_j/k^t_{j,Y2}$ = 35.62. So, in this case, the firms decide to collaborate.

B1.2. Greater Expected Value

This process is associated with the decision made in A2. Firm i decides to collaborate with firm j if the expected growth of the stock of knowledge for firm j, E (k^t_j), is greater than the growth of the stock of knowledge for firm i (Δv_i).

(i) *If* E (k^t_j)>Δv_i *then* cooperate;

(ii) *If* E (k^t_j)≤Δv_i *then* do not cooperate.

B1.3. The concentration process

In the concentration process, the most linked firm is selected as a potential partner. Here, every firm will choose the most linked partner to collaborate with. No other rules interfere in the choice of the partner This is simply the choice of the firm that has more connections with other firms, also known as the “rich gets richer” paradigm of Barabasi (2002).

B1.4. Profit evaluation

This approach was derived from Carayol and Roux (2003) and is based in the profit evaluation of the extended (or reduced) network, that is to say, the profit of the network considering the inclusion or the exclusion of nodes in the network. We denote by **G** the network containing i and j. Through the *profit evaluation* process, a new link between i and j is created if the profit of firm i in case j joins the network, denoted by $\pi^t_i(G_{+j})$, is greater than the profit of firm i in case j doesn't join the network, $\pi^t_i(G)$. Simultaneously, the profit of j in case i joins the network, $\pi^t_j(G_{+i})$, must be greater than the profit of j in case i doesn't join the network, $\pi^t_j(G)$. Therefore, the decision process of creating a link between firms i and j can be formalized as follows:

if $\pi^t_i(G) < \pi^t_i(G_{+j}) \wedge \pi^t_j(G) < \pi_j(G_{+i})$ *then* a link between i and j is created.

Equation 6-29

Along with the same idea, a link between i and j is not created or is deleted (if it already exists) if the profit of i in the current network in case j is in the network is lower than the profit of i in case j is deleted from the network, denoted by $\pi^t_i(G_{-j})$. Simultaneously, the profit of j in case i is in the network must be lower than the profit of j in case i is deleted from the network, that is, $\pi^t_j(G_{-i})$. The decision of not creating a link (or deleting it) is made according to the following rule:

if $\pi^t_i(G) < \pi^t_i(G_{-j})$ and $\pi^t_j(G) < \pi^t_j(G_{-i})$ *then* a link is not created or is deleted in the case where it already exists **Equation 6-30**

We have adapted the original model from Carayol and Roux (2003) by introducing the following restriction: firms which are looking for a partner to cooperate with must have a stock of knowledge value, k^t_j, smaller than the average stock of knowledge of all firms within that market. Table 6-3 illustrates this process with an example from the simulation algorithm.

Cooperating firms		Product		π^t_i (G)	π^t_i (G_{+j})	π^t_j(G)	π^t_j(G_{+i})	k^t_i	E(k^t)
Firm i	Firm j	Firm i	Firm j						
15	1	2	1	5.355	72.484	205.009	212.470	0.581	0.584

Table 6-3– Example of collaborations between firms in Profit evaluation

Note: in this table, E(Kt)=0.584 represents the average value of the stock of knowledge for product Y_2

As we can see in this example, firms 15 and 1 are able to create a link because the profit of i in case j joins the network (or the extended profit of i, $\pi^t_i(G_{+j})$), is greater than the profit of i in case j doesn't join the network, $\pi^t_i(G)$. At the same time, $\pi^t_j(G_{+i})$ is greater than $\pi^t_i(G)$. In this example, the profit of firm 15, $\pi^t_{i=15}(G)$, is computed by ignoring the existence of the firm j. The extended profit $\pi^t_i(G_{+j})$ is computed as follows (with i=15 and j=1):

$$\pi^t_{15}(G_{+1}) = (t_1{}^t - c_1 + \Delta k^t_{15} + \delta^{d^t(15,1)} \cdot w^t_1).y^t_{15} - g(w^t_{15}) \,. \qquad \text{Equation 6-31}$$

Assuming that firm 15 is a seller, this equation is obtained from equations 6-19 to 6-22. The equation of the extended profit of firm 15 (by including firm 1 in its network) is therefore equal as the usual profit, except that expression $\delta^{d^t(15,1)} \cdot w^t_1$ is added to the rest of Δk^t_{15}, (the stock variation), increasing the profit by ($\delta^{d^t(15,1)} \cdot w^t_1$) y^t_{15}.

To delete the link between i and j, we should have $\pi^t_i(G) < \pi^t_i(G_{-j})$ and $\pi^t_j(G)<\pi^t_j(G_{-i})$. In that case, the decreased profit, $\pi^t_i(G_{-j})$, would be computed as follows:

$$\pi^t_{15}(G_{-1}) = (t_1{}^t - c_1 + \Delta k^t_{15} - \delta^{d^t(15,1)} \cdot w^t_1).y^t_{15} - g(w^t_{15}) \,. \qquad \textbf{Equation 6-32}$$

In this case, the expression $\delta^{d^t(15,1)} \cdot w^t_1$ was deleted from the rest of Δk^t_{15}, (the stock variation), decreasing the profit in ($\delta^{d^t(15,1)} \cdot w^t_1$) y^t_{15}. In addition, we can confirm, in this case, that the restriction that we have introduced is satisfied, i.e., that the stock of capital of firm 15, associated with product Y_2 (k^t_i = 0.581), is smaller than the average stock of knowledge for product Y_2, E(k^t) = 0.584.

B2. Collective formation

As we will see later in Section 6.9, some networks are formed as a consequence of a collective will. In these collective initiatives, a group of organizations creates a network by joining at the same time. Collective contracts can exist in these types of networks. Links are created at the same time for all firms that belong to this group. In our simulation we assume that collective networks can be formed just once and can dissolve after the final of the collective contract. One example of such a process of collective formation is the *groups of suppliers* (see Chapter 2) in which the creation of a collective network represent an interesting opportunity for all the cooperating firms to organize more efficiently their supplies.

We will distinguish two types of collective formation of networks: *collaboration networks* in which there is a common goal for the whole network and *cooperative networks* in which a common goal does not explicitly exist.

B2.1 Collaboration networks

In collaboration networks there is a common goal to all members of the network. To form this network, firms agree in setting up a collective contract that lasts for a predefined time interval, during which no new members are accepted.

B2.2. Cooperative networks

Cooperative networks are similar to collaboration networks except that for new members that want to enter the network. In these cases, they must negotiate directly with one of the members of the network via *Stock ratio complementariness* (B1.1). A contract may exist if the parameter *contract* is active in the simulation algorithm.

Virtual Enterprises

We have also introduced the possibility of network formation in which the transactions costs are lower and the physical distance between the elements of the networks has less importance. This is the case of e-Marketplaces, that we can consider as a particular case of a virtual enterprise in which a temporary alliance of enterprises is created. It consists of tasks performed by autonomous teams of small and medium enterprises (Camarinha-Matos and Afsarmanesh, 2004).

To explore this type of relationships we used both types of collective formation: collaboration networks and cooperation networks, because they represent two important models of B2B e-Marketplaces (see Appendix IV).

The main differences between these virtual collective networks and the former collective formation process lay in the pyisical distance which is not taken into account in the choice of the partner. Only the technological distance matters. Therefore, the distance between firms i and j, $d^t(i,j)$**,** is given by equation Equation 6-33,

$$d^t(i,j) = d^{tech,\,t}(i,j), \qquad \textbf{Equation 6-33}$$

where $\mathbf{d^{tech}}$ is given by Equation 6-7**.**

Contracts

The creation of a new link between firms may involve the definition of a contract. In our simulation the existence of a *contract* is configured as a parameter with two possibilities: Contract / No contract. In the case when the contract option has been activated, firms will maintain the link that connects them, until the end of the contract. In the case of no contract, the link between the firms may be deleted when the conditions for the link are no longer justified. For instance, in the negotiation type B1.1, the link will be dropped if stock ratio complementariness is no longer observed.

There are cases where the existence of a contract may not imply the maintenance of a link. When a firm fails, all links connecting the firm that fails to other firms disappear. Nevertheless, to avoid the death of a firm, and the consequent elimination of its links, we may assume a mechanism of collective altruism, that we will call the *Network backup* or *Network support.* Firms that are linked to the failing firm are normally interested in its survival. We will adopt a similar situation in our simulation. This mechanism will be explained in detail later in Section 6.4.8.

6.4.4. Production quantities

In economic literature, every final product producer i must decide about the quantities X_i^t that will produce at time t, having in mind the maximization of the profit. However, as firms have limited information about the demand function and other competitors' production costs, we have used a heuristic approximation to determine their output quantity: $x_i^t = \alpha_i x_i^{t-1}$. We assume that at the beginning of the simulation, firms have null profits. In the following generations, we admit that when the profit of firm i is positive, the output quantities increase by a factor α, which relates with market dynamics:

$$\alpha_i = \left(1 - \varepsilon_i^{t-1}\right)\frac{p_i^{t-1}}{c_i^{t-1}} \quad \textbf{Equation 6-34}$$

where ε_i is the price elasticity of demand[97] of final product i:

$$\varepsilon_i = \frac{\partial x_i}{\partial p_i}.\frac{p_i}{x_i} = -\frac{\sum_{r=1}^{n} A_r.x_r}{\sum_{r=1}^{n} A_r.x_r - b\sum_{r\neq i}^{n} A_r.x_r}$$

We assumed that intermediate good products are combined into a fixed proportion in the production of final good x:

$x = \min(y_1, y_2)$

Therefore, buyers' demand quantity of each intermediate product[98] is:

$y_1 = y_2 = x$

So, from Equation 6-17 and Equation 6-18 and since $y_1 = y_2 = x$, then the prices of inputs are given by:

[97] The price elasticity of demand measures the sensitivity of the quantity demanded by consumers to the price (Besanko and Braeutigam, 2005).

[98] That is, if buyers produce 10 units of final good X, then they will use 10 units of each intermediate product, Y_1 and Y_2.

$$t_1^t = \frac{\sum_{j=1}^{m_1} A_j^t}{\sum_{i=1}^{n} x_i^t}$$

and

$$t_2^t = \frac{\sum_{j=1}^{m_2} A_j^t}{\sum_{i=1}^{n} x_i^t}$$

6.4.5. Migration

In our simulation, firms are placed in two different regions: Region 1, with higher marginal costs and Region 2, with lower marginal costs. Firms are allowed to migrate in order to increase their profit[99].

For now, only firms with negative profits and older than one period may migrate, in order to overcome their bad situation. Migration results will be analysed in Section **6.9**.

6.4.6. Cognitive modelling - Learning, Network support and Risk

As we have presented in Section 3.3, our aim is that the agent simulates diverse cognitive activities, in which we can distinguish *beliefs*, *intentions* and *desires* (the BDI-architecture). The learning component interacts with this mental configuration as all learning relates to beliefs (Ferber, 1999) and desires. In the following paragraphs, we will describe the learning mechanism that was implemented in the architecture of the agents, and some of the attributes that characterize them: one individual attribute, the *risk*, and one collective attribute, named *Network support.*

[99] Only two regions were consdeired, for reasons of model simplification.

Learning

We have implemented a learning mechanism to give to agents the capability of forming certain beliefs of the world where they exist. This *belief* is related with the formation of a network, more specifically with the decision to cooperate with a particular partner.

When links between firms exist, the firms can share information about the most and the least successful partners. A "so called" successful partner should be a *stable* partner, that is, it should be linked for a long time to other firms, in order to maintain cooperation and increase the profit[100]. This attribute of *persistence* of the partner j for a firm i will be defined as $P(i, j)$.

Firm i that may or may not be linked to other firms in the network, can therefore decide to establish a link with a new partner (say, firm j), based on information about the success that firm j has already experienced with other firms. This information exists for all link and includes the identification of the partner j, its level of the stock of knowledge, k^t_j, and its profit, π^t_j.

The learning algorithm that was implemented in agents' architecture is the *instance-based learning*. The algorithm looks for similar related actions retrieved from memory and use them to decide about the new action to adopt. This method is explained in Figure 6-6.

[100] We will see later that firms that cooperate have higher profits, in average. Simultaneously, long-term links increase average duration of life of firms involved.

Instance-based learning algorithm

- *Given a particular pair of firms,* ***a****, including f_i (the cooperating firm) and f_j (the partner with whom is expecting to cooperate):*

  (i) *Identify the total number of pairs* ***p*** *considering all existing firms (f_1, the cooperating firm and f_2, the partner) that established links to other firms in past actions;*

  (ii) *Repeat for all pairs of linked firms,* ***p***:

  - *Compute the average Euclidean distance* ***d(a,p)*** *between the actual pairs (f_i, f_j) and pairs from past actions (f_1, f_2)* using profit and the stock of knowledge:

$$d(a,p) = \frac{\sqrt{(\pi_1^t - \pi_i^t)^2 + (\pi_2 - \pi_j)^2} + \sqrt{(k_1^t - k_i^t)^2 + (k_2^t - k_j^t)^2}}{2}$$

  - *Compute the persistence of the links,* $\hat{P}(f_1, f_2)$, i.e. their average duration.

  (i) *Taking into account the distance* ***d(a,p),*** *the k nearest neighbours* ***p*** *of* ***a*** *will be selected* (a limit is previously defined).

  (ii) *The algorithm returns the average value of the persistence of the links of the k nearest neighbours of* ***a****:*

$$\hat{P}(i,j) \leftarrow \sum_{l=1}^{k} \frac{1}{k} \hat{P}_l(f_{1l}, f_{2l})$$ **Equation 6-35**

Figure 6-6 – Steps of instance-based learning algorithm

The results of the learning mechanism are given in Section 6.9.

Attitude to risk

We have implemented a very simple cognitive model of entrepreneurship[101] to endow agents (firms) with more reactivity. Following Leibenstein (1969), we distinguish two kinds of entrepreneurship: *routine (or conservative)* and *new type (or adventurous). Adventurous* firms establish new connections with other markets, create and expand

[101] Entrepreneurship is the practice of starting new organizations, particularly new businesses generally in response to identified opportunities (often a difficult undertaking, as a majority of new businesses fail).

activities and are attracted to risk in their decisions, while, in general, *conservative* entrepreneurs avoid risk.

For that reason, we have introduced a discrete attribute associated with the *attitude to risk*, in order to control the decisions of the agents involved in the simulation. In our simulation, the *attitude to risk* involves the decisions associated with migration and with avoiding other firms from failure:

(i) *Adventurous* firms migrate to a different region when they are in a difficult situation and

(ii) *Adventurous* firms are determined to save other firms from failure, even if it increases their costs.

An attribute "*risk level*" was used to represent the attitude to risk assigned to each firm. We consider that the risk level takes discrete values within the interval [1, 10]. *Conservative* firms have a negative attitude to risk and therefore are associated to risk levels under 5, while *adventurous* have risk levels greater than or equal to 5. A random risk level is assigned to each firm and remains the same throughout its lifetime.

Network support

In some types of networks, essentially in those where a common goal exists, collective efforts are done in order to save firms from failing. This actually constitutes one of the advantages for firms integrating networks, since intense relationships are established between firms of the same network.

Although in some situations this *network support* could be considered a type of *altruism*, we prefer not to call it that way. Actually, we observe that firms which decide to save other firms (in the same network) from failure are simultaneously helping themselves, since maintaining the network active provides a safeguard or guarantee to the firm in question.

The way how we have implemented this network support in NetOrg is simple: if this mechanism (as well as birth and death parameters) is active, and if any firm, say, A, is about to get extinct due to the predefined conditions (expressed back in Section 6.4.1.), then a collective effort is made to help firm A. These efforts are made by the *adventurous* firms (i.e., firms with a positive attitude to risk) that are connected to A. These efforts consist in transferring some of their knowledge, materialized under the form of stock of capital, to the failing firm.

In the next sections, we focus on technical aspects of the simulation algorithm, namely the model evaluation, generic issues about communication between agents, the configuration of model parameters, implementation and issues concerning software for MAS. In Section **6.9** the results are provided.

6.5. Communication between agents

Communication between agents is of great importance in Agent-Based simulation. Our framework does not include protocols as KQML or ACL that were referred to in Section 3.3. However, we define other levels of communication that motivate individual and collective actions of agents. These levels do not take into account any formal protocol.

- **Group communication**

  Agents react to the presence of other agents and to the conditions of the environment; *migration* is an action driven by rules used by agents when evaluating the environment within the *beliefs*, *desires* and *intentions* architecture; Network Support (a kind of an altruistic mechanism) is a collective action that is undertaken in order to save a partner firm from failure.

- **Individual (peer-to-peer communication)**

  Agents search for the best partner for cooperation; in this search the agent takes into account the distance between both and the type of partner. For an agent to be aware of others and to be able to take this kind of decisions, we endow the agents with individual communication capability.

6.6. Simulation algorithm

In this section, the simulation algorithm is presented both using an outline of the pseudo-code and a flowchart.

Pseudo-code of NetOrg (outline)

1. Initialization

1.1. Initialization of general parameters

1.2. Definition of the number of firms and consumers

1.3. Creation of firms

1.3.1. Firms are placed in the grid

2. Cycle *(iteration cycle for a fixed number of generations)*

2.1. Knowledge and financial indicators and cooperation decision

2.1.1. Innovation and knowledge accumulation

2.1.2. Market Demand and other financial indicators

2.1.3. Cooperation (according to a predefined strategy)

2.2. Migration

2.3. Production Quantity and Age

2.4. Birth and Death

2.4.1. Death

2.4.2. Birth

2.4.2.1. Location

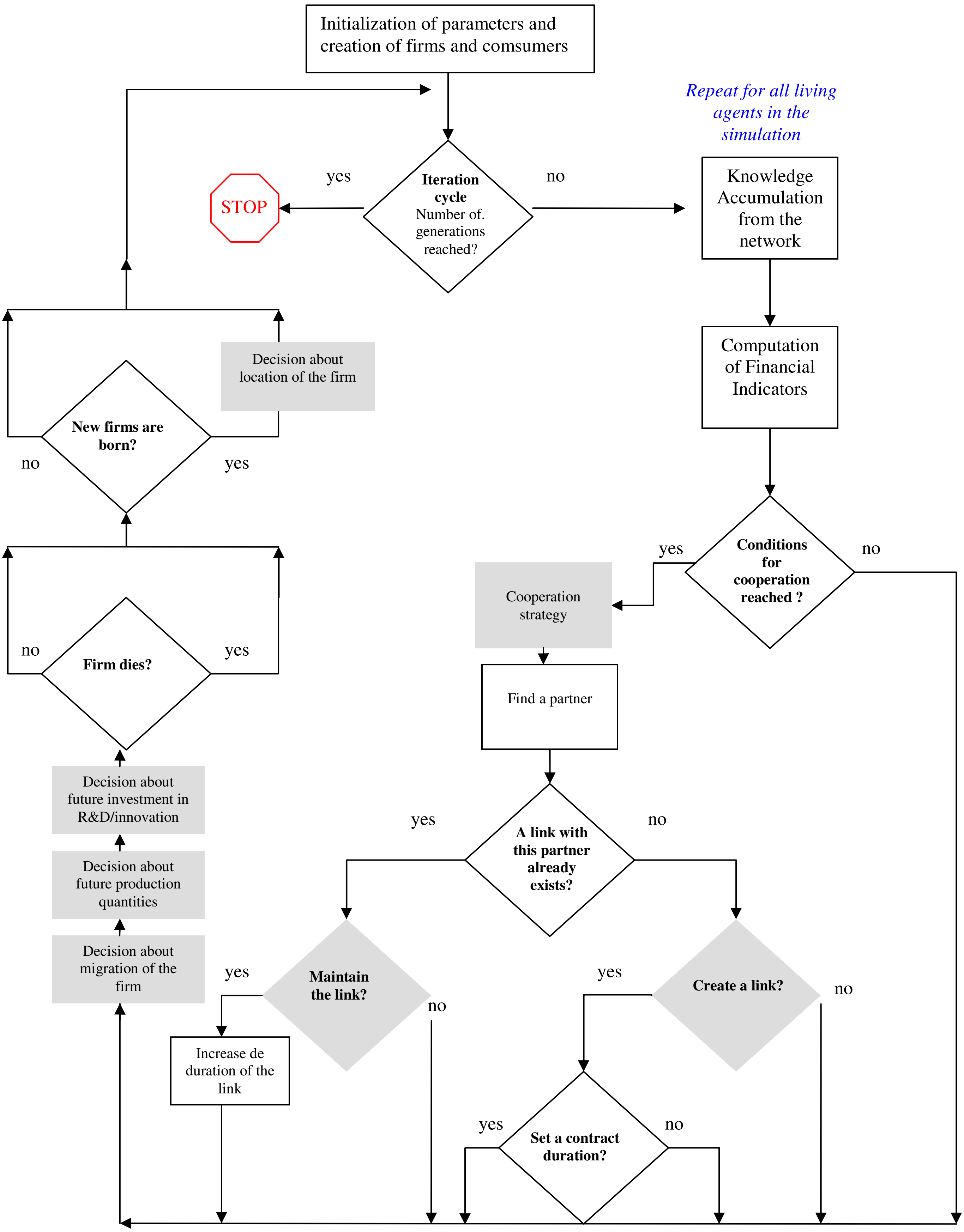


Figure 6-7 – Flowchart of the NetOrg simulation algorithm. Steps in shaded boxes are associated with the cognitive model. The algorithm runs for 10 or 20 generations (iterations).

6.7. Implementation and Software for MAS

Our simulation has been implemented using R, a language and environment for statistical computing and graphics (R Development Core Team, 2005). R can be considered as a different and free implementation of S which was developed at Bell Laboratories (formerly AT&T, now Lucent Technologies) by John Chambers and colleagues (Becker and Chambers, 1985). Some important differences between S and R exist, but much code written for S runs unaltered under R.

There are, of course, specific languages for MAS that could have been used in this work[102], but we preferred to use R as various statistical computations are needed and so we exploited various functions from statistical libraries[103].

The simulation contains the following programs and functions created by the author:

- *NetOrg, v. 4.7* (the main program), which runs the whole simulation and includes the following functions:
  - `Firmdeath`: determines if death occurs according to the survival probability defined under the density dependence model;
  - `location`: determines the place (coordinates) in the space for a newborn firm;
  - `LearningCooperationDecision`: defines whether the firm cooperates according to instance-based learning;
  - `DecisionPartners`: defines whether the firm that is about to get extinct can be rescued by the Network Support mechanism;
  - `generateGraph`[104]: depicts graphs, according to the adjacency matrix.

---

[102] Examples of specific software and frameworks for MAS are JADE, Swarm, Ascape, Netlogo, etc.

[103] One of the most useful R libraries that we use in NetOrg is the *Igraph library* developed by Gabor Csardi, Csardi (2005) from the KFKI University in Budapest. *IGraph* contains functions for generating regular and random graphs according to known algorithms and models in the network theory literature. A set of structural property calculation functions like diameter, transitivity, etc. are also included.

[104] We are grateful to Kaustubh R. Patil for helping us to with the implementation of this function. It needs the library *Rgraphviz,* which interfaces R with the AT&T GraphViz library to provide the ability for plotting R graph objects from the graph package.

Besides, *NetOrg* also calls another program (`Network Analysis for NetOrg`), in order to organize the information about the networks. In fact, it is a difficult task to follow which network is which because they are dynamic entities that are continuously changing their shape[105].

To help in the organization of the output data for specific purposes, as survival analysis and some statistical tests, other functions have been created by the author, as `Survival` and `LinkStability`. In the Appendix III we include function `Firmdeath` as an example of a function that has been used in NetOrg.

6.8. Identification of the hypotheses and model validation

Considering the main observations that were identified earlier in this work considering three different industries, we have built a table that systematizes all this information and makes a correspondence with the general questions of Chapter 4.

[105] In some cases it was too difficult to follow the evolution of networks and to distinguish which network is which. For instance, if a network breaks apart, producing two different networks in the following generation, then we consider that the network keeping the majority of the nodes of the breaking network is the same network that broke apart. Therefore it keeps the same identification. If two networks merge into one only, then we consider that the resulting network maintains the identification of the network that contains the most part of the nodes (between the two merging networks).

Specific Hypothesis	Observations/Evidences obtained from literature	Test applied	Correspondence with general questions presented in Table 4-1
H_{21}	Buyers promote both a concentration of suppliers and room for creating important collaborations and alliances with suppliers and among suppliers and buyers; Networks are formed and several topologies are observed.	Empirical observation	
H_{22}	Some collaborative firms have increased their profits because of the reduction of costs due to the rationalization of some of their functions [through collaboration via network formation].	Mann-Whitney Test	G4
H_{23}	Migration of firms to markets with lower marginal costs is as a way to rationalize production	Empirical observation	
H_{24}	Contemporaneous density has a negative impact on the mortality of organizations	Survival Analysis	G1
H_{25}	Density at founding has a positive impact on the mortality of organizations	Survival Analysis	G1
H_{26}	Contemporaneous density (of networks) has a negative impact on the mortality of the networks	Survival Analysis	G2
H_{27}	Density at founding (of networks) has a positive impact on the mortality of the networks	Survival Analysis	G2
H_{210}	Contemporaneous density has a negative impact on the mortality of networks	Survival Analysis	G3
H_{211}	Density at founding has a positive impact on the mortality of networks	Survival Analysis	G3
H_{28}	Firms that have long-term relationships with other firms will live longer, in average	Mann-Whitney Test	G7
H_{29}	If the distance between two firms is short, they will have a higher tendency to cooperate horizontally; if the distance between two firms is high, they will have a higher tendency to cooperate vertically	Mann-Whitney Test	G8

Table 6-4 – Summary of the Hypothesis that will be tested with NetOrg

Hypothesis H_{24}, H_{25}, H_{26}, H_{27} and H_{210} and H_{211} are similar in what concerns the type of analysis. However, the levels of analysis are distinct: while H_{24}, H_{25}, H_{26}, H_{27} are tested considering the same level of analysis (only firms or only networks), hypotheses H_{210} and H_{211} measure the impact of firms in the survival analysis of networks.

These hypotheses are going to be tested as a part of the process of model validation. As presented in Section 3.3.3, validation is important as it provides a way of obtaining the credibility of the simulation model. Our research methodology presents a practical way to evaluate the model by comparing simulation outputs and empirical studies in order to verify that the model is able to reproduce some historical data of the target industry.

In *NetOrg*, validation of the model includes the following steps:

(i) Analyse the impact of different parameter values in the simulation; this step is important to define the initial values for simulation parameters and corresponds to the *stage 1* of the *two stage* research methodology presented in Figure 4.3. This step is carried out in Section 6.9.1.

(ii) Run the model for different cooperation strategies and analyse if there is consistency between simulation outputs and empirical studies. The testing of some hypotheses that reproduce reality (presented above in table 6.4) can be used to analyse this consistency. This is done in Sections 6.9.2, 6.9.3. and 6.9.4

(iii) Explore more general questions about cooperation networks and industrial evolution, as the performance of the strategies, the evolution of network indicators, and the identification of evolutionary patterns. To achieve this goal we used Data Mining techniques as decision trees, and Multivariate Data Analysis Techniques presented in Sections 6.9.2 and 6.9.5.

Our aim is to carry out validation and check whether all these three steps have been satisfied.

6.9. Experimental results

NetOrg is a Multi-Agent model that has been used to capture the survival as well as the general dynamics of the network formation. The results given in this section are the outcome of the average of more than 100 different runs of the simulation where, in some cases, we have taken one representative output to illustrate the evolution of the graphs representing networks of firms.

We start by defining and initializing all the parameters that are necessary to run the simulation. After that, we analyse and compare the collaboration strategies that reproduce networks' literature. In this analysis we use non parametrical tests and decision trees to compare the performance of networks and survival analysis (through Cox Regression) to explore the impact of some variables on the mortality of organizations and networks. Finally, multivariate techniques are exploited to analyse the relevance of the variables on the evolution of the organizational networks.

6.9.1. Model parameters

Configuration

We have defined a set of parameters considering that industry X would represent a buyer and industries Y_1 and Y_2 its suppliers. The marginal cost was defined differently for the three types of industries: marginal costs in region 2 were half of the corresponding values in the region 1.

In the following tables, we present the initial configuration of parameters. Table 6-5 presents the user defined parameters for the economic model, and Table 6-6 contains the user defined parameters for the general simulation set up.

User defined parameters for the economic model	Description	Notes	Values
N_X	Number of consumers of the final market X		N_x=5000
R_{Xi} and uR_{Xi}	Income of the individual consumer, i, (=1, 2, …, Nx) of the final market X and update value for the income of consumer i	The income of consumers in market X is randomly defined by an Uniform distribution between a minimum and a maximum level. In every iteration, the income of a particular consumer is updated by a factor randomly defined (also by an Uniform distribution) in the following way: $R_{Xi} = R_{Xi} x\, uR_{Xi}$	R_{Xi}~U[50 ; 200] uR_{Xi}~U[-1 ; 1]
P_{ki}	Price of capital	Price of capital for each of the three markets	Pk_X=3 Pk_{Y1}=3 Pk_{Y2}=3
k_i	Stock of Knowledge for the firms in each of the three different markets (X, Y_1, Y_2). Initialization of k_i is made with w^t_i, the innovation at instant t.	To compute the number of firms in the beginning of the simulation, the average stock of knowledge is used. This average stock is the mid point between the minimum and the maximum stock defined previsouly for each the initial values of the Stock of Knowledge in each market (k_{iX}, k_{iY1}, k_{iY2}).	$Min(k_{iX})$=1 $Max(k_{iX})$=20 $Min(k_{iY1})$=1 $Max(k_{iY1})$=15 $Min(k_{iY2})$=1 $Max(k_{iY2})$=15
DS_u, DS_l	Density Survival interval		DS_l=1 DS_u=14
DB_u, DB_l	Density Birth interval		DB_l=3 DB_u=5
S_l	Size limit		4
A_l	Age limit		2
X_1, Y_1, Y_2	Quantity of product produced by firm i	Initial quantity is obtained from a Uniform distribution.	X~ U[1;500] Y_1~U[1;500] Y_2~U[1;500]
α_i	Factors that increase or decreases the production quantity of a firm	Quantity is updated in every iteration using a heuristic for the factor α that increases or decreases the production quantity, based on the profit of a firm	see Section 6.4.4
b	Indicator of the association between the final good varieties; b=1 means that goods are perfect substitutes		b=0.7
δ	Transferability factor	Parameter that measures the proportion of new knowledge which is effectively transmitted through each edge.	δ=0,9
MC_X MC_{Y1} MC_{Y2}	Marginal Cost of the firms in each of the three different markets	The marginal cost for a particular firm is defined randomly based on a Normal Distribution with mean = MC_i (i.e. the marginal cost in one particular market i, with i=X, Y_1, Y_2) and standard deviation = 0.1Xmean	MC_X=0.1 MC_{Y1}=0.05 MC_{Y2}=0.05
w^t_i	Innovation of firm i at time t	Innovation w^t_i is a function of the knowledge previously owned by firm i with a certain probability, p, given by k_i^{t-1} p, where p is based on a Normal Distribution with mean λ and standard deviation = 0.1X λ. We assume that λ differs according to each market (X,Y_1,Y_2), the value of λ_X being greater since it corresponds to buyer (larger firms)	λ_X=0.03 λ_{Y1}=0.01 λ_{Y2}=0.01
[a',b']	Limits for firms to be considered similar in terms o technological distance		[5, 1000]

Table 6-5 - User defined parameters for the economic model

User defined parameters for simulation's set up	Description	Notes	Range	Default value
Birth	Activates the possibility of births of firms	0=no births; 1=births	{0,1}	0
Death	Activates the possibility of death of firms	0=no deaths; 1=deaths	{0,1}	0
Migration	Activates the possibility of migration of firms	0=no migrations; 1=migrations	{0,1}	1
Contract	Activates the contract between firms	0=no contracts; 1=contracts	{0,1}	1
Strategy	Defines the strategy of cooperation		{1,2,3,4,5,6,7,8}	
Learning	Activates the possibility of learning	0=no learning; 1=learning	{0,1}	1
Generations	Number of generations in the simulation	Usually defined as 20	{10, 15, 20}	20
Runs	Number of runs	Usually defined as 10	{5, 10, 15, 20}	10
Cooperation	Activates the cooperation between firms	0=no cooperation, 1=cooperation	{0,1}	1
Network Support	Activates the network support	0=no network support; 1=network support	{0,1}	0
Neighboughrood Ring (Neighbourhood Diameter)	This parameter establishes a circle of radius 1, 2 or 3 cells which defines the diameter of the circle (or ring) of neighbours around a particular firm.		{1, 2, 3}	3

Table 6-6 - User defined parameters for simulation's set up

All parameters presented in Table 6-5 and Table 6-6 are defined previously as they do not depend on others. We will see that in some cases, initializations depend on other parameters of the simulations.

Initialization of the simulation

As said before, no fitness measures are used here to adjust the parameters of the simulation. Corresponding to *stage I* in the research methodology (Figure 4-3), the choice of the initial values for the parameters was defined in such a way that the simulation would present reasonable results. Thus, after running the simulation for several times, we have found some problems with the magnitude of the variables. For instance, the profit of the firm was reaching very high values (or very low) when compared to other variables. At the same time, we observed the complete extinction of firms in few generations or else a "*demographic explosion*". For that reason, we made some changes in the initial settings of parameters. Table 6-5 presents the choices we made to reproduce reasonable patterns of behaviour. In other words, the choice of parameters was made in order to produce an output corresponding to the following goals:

- *Limited variance of some variables*: the coefficient of variation of some variables that follow the Normal distribution was set in 10%. For instance, innovation, w^t_i, depends on p, which is a random variable that represents the update of innovation, being normally distributed with mean λ and standard deviation = 0.1X λ.
- *Coherent processing time and simplicity of the ouput*: the number of firms should not be very small neither very large. Few firms do not produce enough number of networks to be analysed. A large number firms slows the algorithm, since the processing time grows drastically with be number of firms.

As we stated before, reasonability is a somewhat subjective concept, but the idea here is to construct intervals of tolerance for the parameters in such a way that the results produced by the parameters belonging to those intervals are realistic enough to be conserved. Besides, *Stage I* of our research methodology (Figure 4-3) does not end with this initialization. In fact, some adjustments in the initial parameters were done after *Stage II*. We repeated these two states for several times, in order to obtain consistency between the simulation output and the facts from empirical studies[106].

The initialization of the number of firms in each market is obtained randomly, although it was somewhat controlled, based on the initial values of the knowledge stock in each market:

$$n= f(K_x);\ m_1= f(K_{Y1});\ m_2= f(K_{Y2}) \qquad \text{Equation 6-36}$$

[106] Empirical studies are used as an orientation of our simulation. As explained above, there is no great concern in that the simulation follows exactly the same properties of these empirical studies.

with $\mathbf{N=n+m_1+m_2}$, and **n,** $\mathbf{m_1}$ and $\mathbf{m_2}$ being respectively the number of firms in markets **X,** $\mathbf{Y_1}$ and $\mathbf{Y_2}$. The function **f** was defined as depending on K^t_i (the stock of knowledge), and R_X, R_{Y1} and R_{Y2}, (consumers' average income in each market) as follows:

$$n = \frac{E(R_x)}{P_{k_X} \frac{Min(k_{iX}) + Max(k_{iX})}{2}}$$ **Equation 6-37**

$$m_1 = \frac{nP_x}{P_{kY_1} \frac{Min(k_{iY_1}) + Max(k_{iY_1})}{2}}$$ **Equation 6-38**

$$m_2 = \frac{nP_x}{P_{kY_2} \frac{Min(k_{iY_2}) + Max(k_{iY_2})}{2}}$$ **Equation 6-39**

In these equations, P_k represents the price of capital and P_x represents the price of the final good and it was defined in Equation 6-14. For a numerical example corresponding a particular simulation, given the initial values for $Min(k_{iX})$, $Max(k_{iX})$, $P_{k,X}$, etc., presented in Table 6-5, we obtained $E(R_X)=320.72$, $E(R_{Y1})=572.49$, and $E(R_{Y2})= 785.75$, etc. Therefore, the number of firms in each market was $n=3$ $+m_1=9$, $m_2=18$ (the total being $N=30$).

The price of the capital (P_k) was set to 3 in all of the industries as it seemed important not to discriminate capital values in different markets. Marginal cost for firms of industry X was set to 0.1 and marginal costs for firms of industries Y_1 and Y_2 were equal to 0.05. This difference is due to the fact that companies from industry X represent buyers – that are usually associated with higher marginal costs. The values of the innovation probabilities, (w_i^t) were taken from the Normal distribution with mean 0.03 for industry X and with mean 0.01 for industries Y_1 and Y_2. Corresponding standard deviations were set up to 10% of their mean values.

The values of the density parameters, size and age (DS_l, DS_u, DB_l, DB_u, S_l, A_l) were obtained randomly within the parameter intervals defined in Chapter 5, and are defined in Equation 6-40:

$<DS_l, DS_u, DB_l, DB_u, S_l, A_l>=<1, 14, 3, 5, 4, 2>$ **Equation 6-40**

This set of parameters is more tolerant in what concerns the survival than for new births[107]. Finally, the *Neighbourhood ring* was set in 3. This was also defined randomly among the possible values of this parameter ({1, 2, 3}).

Considering the simulation set up, the majority of the results were obtained given that new firms could not get born nor die. The reason for this choice is that the analysis of network evolution gets very complex if the parameter *death* is activated. Therefore, the parameters *birth* and *death* were disabled (turned to zero) in the majority of the simulation runs. *Network support* was also turned to zero, except when it was used, and *learning* was always activated (turned to 1 in all runs)[108]. *Contract* has also been activated (Contract=1).

The main variables of analysis that have been captured are: profit (π_i), stock of knowledge (k_i), marginal cost, (c,) form, and some other variable concerning network statistics (number of networks, nodes, etc.), at the firm or network level[109]. These variables were collected within each strategy.

Results provided in the following sections are the outcome of more than 100 different runs of the simulation, from where we have taken some representative outputs. The number of initial firms depends on Equations 6-37 to 6-39 and is defined endogenously at the beginning of NetOrg. Usually it ranges from 20 to 40 firms initially, according to the parameterization described earlier.

The simulation results are shown separately: in Section 6.9.2, we give some network statistics, according to different collaborating strategies and for different values of user defined parameters. We test some hypotheses in Section 6.9.3, and in Section 6.9.4 we proceed with the validation of the model using Survival Analysis. In Section 6.9.5.we

[107] We recall that the solution obtained in Chapter 5 was: $<DS_l, DS_u, DB_l, DB_u, S_l, A_l>=<5, 28, 2, 20, 2, 1>$ and was optimized for the region of Ave.
[108] In the analysis that follows, some exceptions on these considerations were introduced. For instance, although we opted to disable *death* and *birth* in the majority of the runs, these parameters are activated in order to analyse the survival at the firm level (results are presented in Section 6.9.4).
[109] Average values of profit, stock of knowledge, and marginal cost were considered at the network level.

analyse the evolution of networks with evolutionary data analysis and, finally, in Section in 6.10, the concluding remarks are drawn.

6.9.2. The dynamics of the network formation

In Section 6.4.3., we have presented the processes that have been considered for the formation of the networks. Two phases corresponding to two decision processes, in which firms can link together to form a network, were considered: *selection of the partner* and *network formation strategies*. Those phases contain several alternatives as we have considered in that section:

A. Selection of the partner:

- A1. Preferential meeting process
- A2. Average stock scenario

B. Network formation:

- B1. Individual decision for integration in new/previous network
  - B1.1. Stock ratio complementarity
  - B1.2. Great Expected Value
  - B1.3. Concentration process
  - B1.4. Profit evaluation
- B2. Collective formation
  - B2.1 Collaboration networks
  - B2.2. Cooperative networks

Analysis of collaboration strategies

In order to analyse the performance of the networks, we have combined those phases and created six collaboration strategies for *selection of the partner* and *network formation* processes, as follows:

a) *Peer-to-peer complementariness (Preferential meeting process,* **(A1)** with *Stock ratio complementariness* ***(B1.1))***
b) *Average Stock Scenario (***A2.***)* with *Great Expected Value* **(B1.2)**
c) *Preferential meeting process* **(A1)** with *Profit evaluation* **(B1.4)**
d) *Concentration process (***B1.3***)*
e) *Collaboration networks (***B2.1+a***)*
f) *Cooperation networks (***B2.2+a***)*

Two more strategies are analysed afterwards. They are associated with the evidences of e-Marketplaces in which the transactions costs are reduced and the physical distance between the elements of the network has no importance. These strategies are identified as:

g) *Virtual collaboration networks (***B2.1+a***)*
h) *Virtual cooperation networks (***B2.2+a***)*

In terms of network formation, strategy g) has the same properties of collaboration networks and strategy h) is similar to cooperation networks. The main differences are in the physical distance, which is not taken into account in the choice of the partner. Only the technological distance matters[110].

The creation of these strategies reproduces what exists in literature and also in empirical evidence. Strategy *a)* is a mix of the works by Carayol and Roux (2003) and by Campos, *et al.* (2006). Strategy *b)* is a common strategy tested in game theory models. Strategy c) uses the *preferential meeting process* and includes the idea of profit evaluation as in Carayol and Roux (2003). Strategies *d)*, *e)* and *f)* are based on empirical observations by CENESTAP (2000a to 2000e, 2001). Finally, strategies g) and h) are associated with virtual enterprises inspired in the observations of e-Marketplaces.

Now, we are going to compare the networks resulting from each of these collaboration strategies. We aim at comparing the performance of the strategies using some economical and network indicators. Results show the network evolution and include the

[110] It is important to note that despite the large number of firms that are typical from these types of virtual networks, as observed in empirical evidences (see Appendix IV), we will not be able to simulate networks with the same high dimension, as it implicates a computation time that is not viable. Therefore, the output will not reproduce the dimensional properties of virtual enterprises.

analysis of profit, marginal cost, average stock of knowledge and several network statistics.

Graphs are represented with the purpose of illustrating the evolution of the networks[111]. These graphs are representative examples of the networks produced in different runs of NetOrg.

a) ***Peer-to-peer complementariness*** *(Preferential meeting process, (A1) with Stock ratio complementariness (B1.1))*

In strategy a) *Preferential meeting process,* **(A1)** with *Stock ratio complementarity* ***(B1.1),*** identified as *Peer-to-peer complementariness,* we have defined a meeting probability, $m^t_{i,j}$, between two firms, in which firm i chooses firm j that maximizes the probability of the similarity between these two firms (that value being equal to $m^t_{i,j}$=[1-d^t(i,j)]).

Usually there is no network formation in the first three to five generations because firms have not enough stock of knowledge to start a cooperation process. When the simulation starts, firms' stocks of knowledge are small and network formation is rare. Along their lifetime, the stock of knowledge of each firm increases and they start negotiating for cooperation purposes[112]. Table 6-7 shows an example coming out from the simulation with the negotiation results that allowed the first cooperation between three pairs of firms based in the *Stock ratio complementarity's* indicator, π^t_i / $K^t_{i,\cdot}$ (in which "·" designates the Market, X, Y_1 or Y_2)

[111] Several topologies were observed (linear, star and multipolar), but the we noticed that the majority of the networks are star and multipolar networks (according to the definitions presented in Chapter 2).
[112] To form the network via the *Stock ratio complementarity* process, we have imposed some restrictions in terms of horizontal and vertical collaboration. Therefore, in order to simulate most part of the supplier/buyer processes, firms only select other firms within the same market (horizontal collaboration between suppliers) or else they select firms of type X (buyers).

	Cooperating firms		Product/ Production Process		Profit / Knowledge Stock				Region	
#	Firm i	Firm j	Firm i	Firm j	$\pi^t_i / k^t_{i,Y=1}$	$\pi^t_j / k^t_{i,Y=1}$	$\pi^t_i / k^t_{i,Y=2}$	$\pi^t_j / k^t_{i,Y=2}$	Firm i	Firm j
1	3	5	1	2	17.88	8.79	82.51	87.02	1	1
2	22	3	3	1	26.24	27.07	659.72	289.43	1	1

Table 6-7 - First cooperation results of NetOrg based in the strategy a)

The corresponding networks formed during this process of partners' selection and negotiation throughout the generations are represented in Figure 6-8[113]:

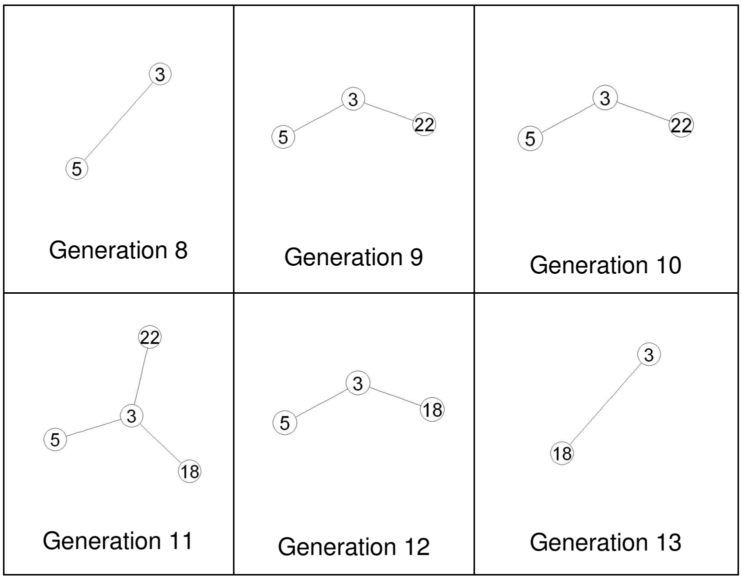
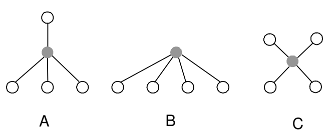


Figure 6-8 – Graphical evolution of the networks created via strategy a)

As it can be observed, networks are not abundant is this simulation[114]. In fact, only few networks were created during the twenty generations. This seems to be a common situation in some cases of network formation (as referred to in the several works by CENESTAP [2000a to 2000e, 2001]). The first network was formed at generation eight

[113] In this case and in further examples of representations of networks, we consider that no hierarchy exists among the firms. For that reason, the following examples of network forms A, B, and C, in which one dark node is linked to four white nodes, are considered equivalent:



[114] Networks emerge at generation 3 or later, because in prevous generations the levels of profit , stock of knowledge, etc. are lower and it is not possible to start any type of negotiation between firms.

involving firms with numbers 3 and 5. Table 6-7 shows that this network was formed because firms 3 and 5 had complementary levels of their stocks of knowledge. At generation nine firm 22 entered the network by linking with firm 3. This network has kept its shape and size during the following two generations and, at generation eleven, firm 18 links with firm 3. In the remaining generations, the links break and the network disappears. In this case, there are two main reasons for networks to disappear. It can be that there is no more stock ratio complementariness between the firms in the network, or else the contract between them is no longer active. These two reasons may occur simultaneously.

b) Average Stock Scenario (A2) with Great Expected Value (B1.2)

Using this strategy, firms first try to find a recognized partner that is already linked with other firms and compute its expected growth to see if the stock of capital is increasing. If the stock of knowledge of the partner is expected to grow, then a new link is established (we therefore assume that, in these conditions, the partner will agree with the link). Otherwise the link will not be established. Table 6-8 shows an example that leads to the creation of a link between firms 1 and 32 because the expected stock of capital, $E(K^t_j)$ of market Y_2 is greater than the growth rate (Δv_j) of the same market.

Cooperating firms		**Product**		**Expected stock of knowledge $E(k^t_j)$**	**Growth rate $\Delta v_j = \frac{k_j^{t-1}}{k_j^{t-2}} - 1$**
Firm i	Firm j	Firm i	Firm j		
1	32	1	3	0.570	0.356

Table 6-8 – Example of cooperation results of NetOrg based in the *strategy b)*

In Figure 6-9, we can see an example in which the shape of networks evolve according to this strategy.

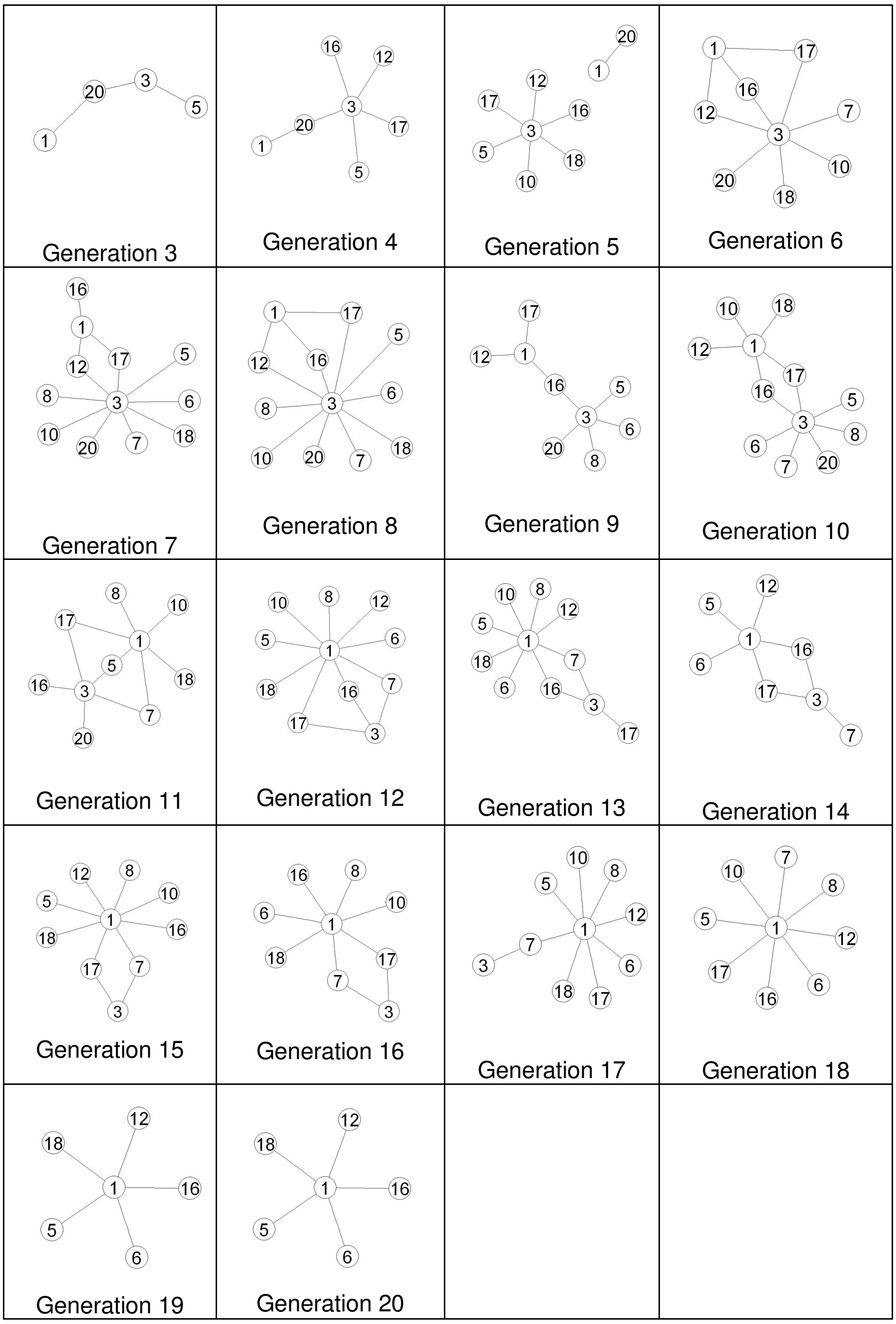


Figure 6-9 - Graphical evolution of the networks created via strategy b)

Networks are denser in this strategy if we compare it with the earlier. Arising at generation three, first network includes firms 1, 20, 3, and 5. At generation four this network expands and incorporates firms 12, 16 and 17 (firm 3 becoming the center of the network). After some changes two centers emerge in this network (firms 3 and 1), which is easy to recognize on the graph of the eleventh generation (see Figure 6-9). There is one whole network formed by two clusters: a group centred at firm 1 (a buyer) and another group centred at firm 3 (also a buyer). Then, firm 3 disconnects and the network structure becomes almost stable: a star vertical network that is centered in firm 1.

c) ***Preferential meeting process*** **(A1) with** ***Profit evaluation*** **(B1.4)** [115]

As we have stated before in Section 6.3.2, a new link between two firms i and j is created within this strategy if $\pi^t_i\ (G_t)<\pi^t_i\ (G_{t+j})$ and $\pi^t_j(G_t)<Pj(G_{t+i})$, where G is the network where i and j belong and π^t its profit at time t. A link between i and j is not created or is deleted (if it already exists) if $\pi^t_i(G)<\pi^t_i(G_{-j})$ and $\pi^t_j(G)<\pi_j(G_{-i})$. Although we based this proposal on Carayol and Roux (2003), an additional restriction has been introduced: firms which are looking for a partner to cooperate with must have their stock of capital smaller than the average stock of the capital. Figure 6-10 exemplifies the evolution of the networks according to this strategy:

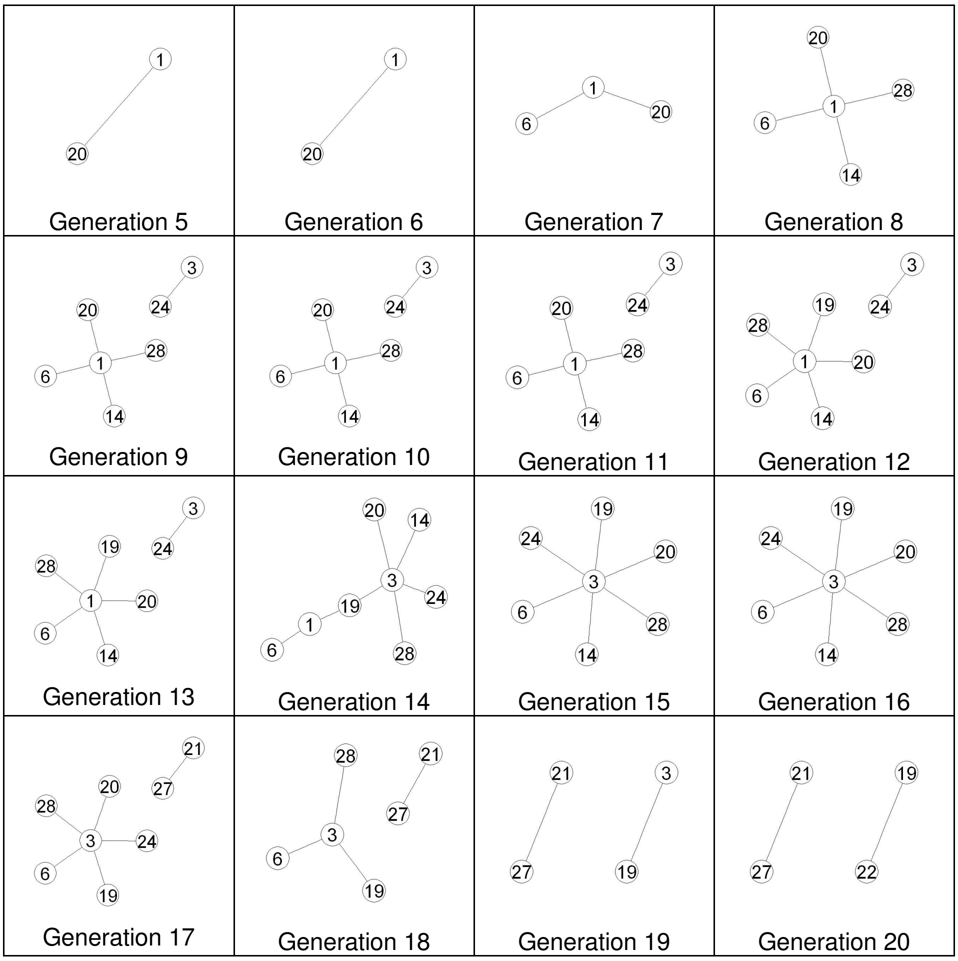


Figure 6-10 - Graphical evolution of the networks created via strategy c)

[115] We have presented a numerical example of this networking strategy back in Section 6.3.2 to illustrate how this process works.

After few generations, a star network grows and at generation eigth, includes firms 1, 6, 14, 18, and 28 (firm 1 being the center of the network). At the same time, firms 3 and 24 share a link and then, at generation 14, integrate the larger network. Firm 3 turns into the new center of the network until generation 19 in which the majority of the links in the network break down. A new network linking firms 21 and 27 emerges at generation 17. This network lasts until the end of the period of analysis, together with a small network joining firms 19 and 22.

d) Concentration process (B1.3)

In the concentration process, firms just look for the most linked firm to cooperate with. No previous negotiation or complex process of partner selection exists. This is simply the choice of the firm that has more connections with other firms. At the beginning, *sotck ratio complementatiness* is used, but when some links exist, firms choose the partner with more links. Figure 6-11 shows an example of such a concentration process where some firms start linking to firm 2 (a buyer) without delay at generation three.

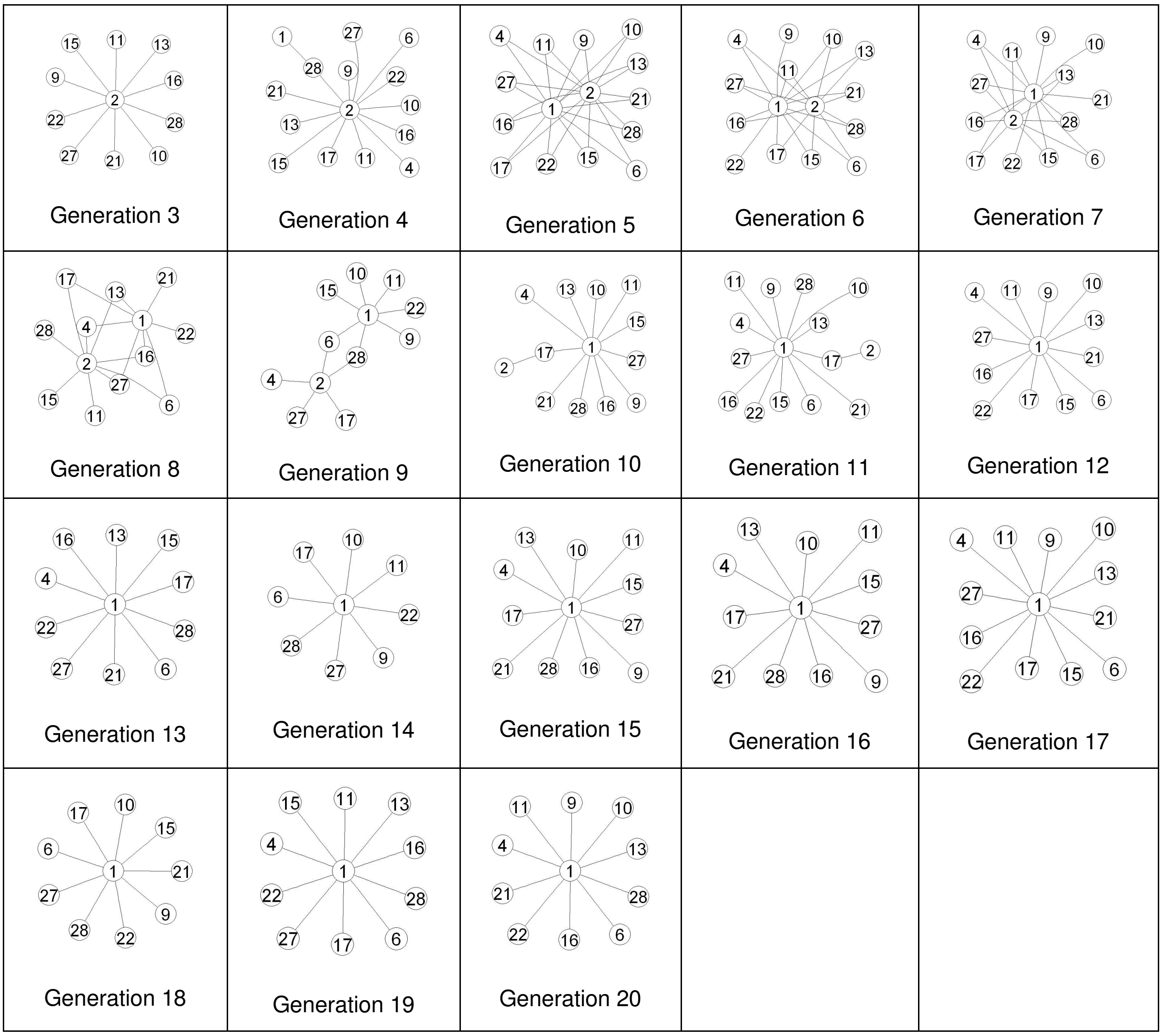


Figure 6-11 - Graphical evolution of the networks created via strategy d)

When firm 1 (a buyer) enters the network, at generation four, the other firms join up and a great number of links arise at generations five to seven. During those generations the graph is difficult to depict since some links cross the others, although it is possible to presume (and it was confirmed) that they are oriented towards nodes 1 and 2. After some generations, the separation of the two subgroups is clearer despite the fact that they are in the same network. Firm 2 leaves the network at generation twelve and firm 1 keeps being the center of the network in the remaining generations.

e) Collaboration networks (B2.1+a)

In collective formation networks, links are created at the same time for all the firms within a group. In collaboration networks there is a goal which is common to all firms of a networks and we assume that they will not admit new elements throughout their existence. Other firms outside this network use a different strategy, *a) Peer-to-peer complementariness,* to select the partner. Figure 6-12 shows an example of such type of networks:

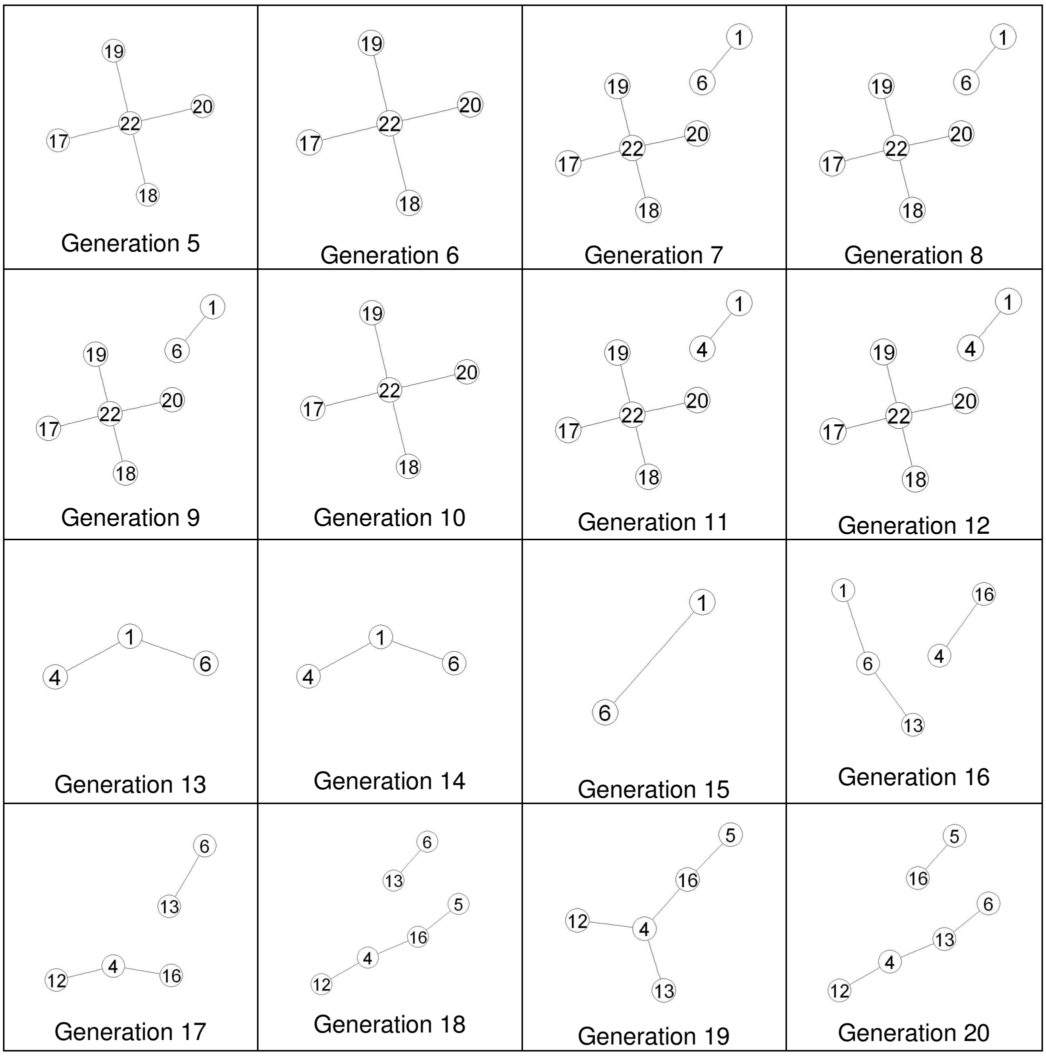


Figure 6-12 - Graphical evolution of the networks created via strategy e)

A set of five firms (identified by the numbers 17, 18, 19, 20, and 22) formed a collaboration network at generation 5. As it can be seen from the evolution of the networks, according to this strategy, this network lasted for eight periods until generation 12[116]. At the same time, at generation 7 another network was created linking firms 1 and 6. This is not a collaboration network (it changes its shape during the lifetime and disappears at generation 10, coming later with a different shape).

In other examples we have observed that some of the firms of the collaboration network keep together in cooperation, even after the contract ends. This may suggest that firms experiencing collaboration networks can gain in terms of stock of knowledge that will be used in further collaborations with other (or the same) firms. In fact, those two firms maintain their collaboration until the last generation.

f) Cooperation networks (**B2.2+a*****)***

As proposed previously in Section 6.4.3, cooperation networks are similar to collaboration networks except that for new elements that want to enter the network. In these cases, they must negotiate directly with one of the elements of the network. So, the common negotiation type will be the *stock ratio complementariness* (B1.1).

[116] In collaboration networks we assume that firms agree in setting up a collective contract that lasts for a predefined interval of time. No new members are accepted within the period of this contract.

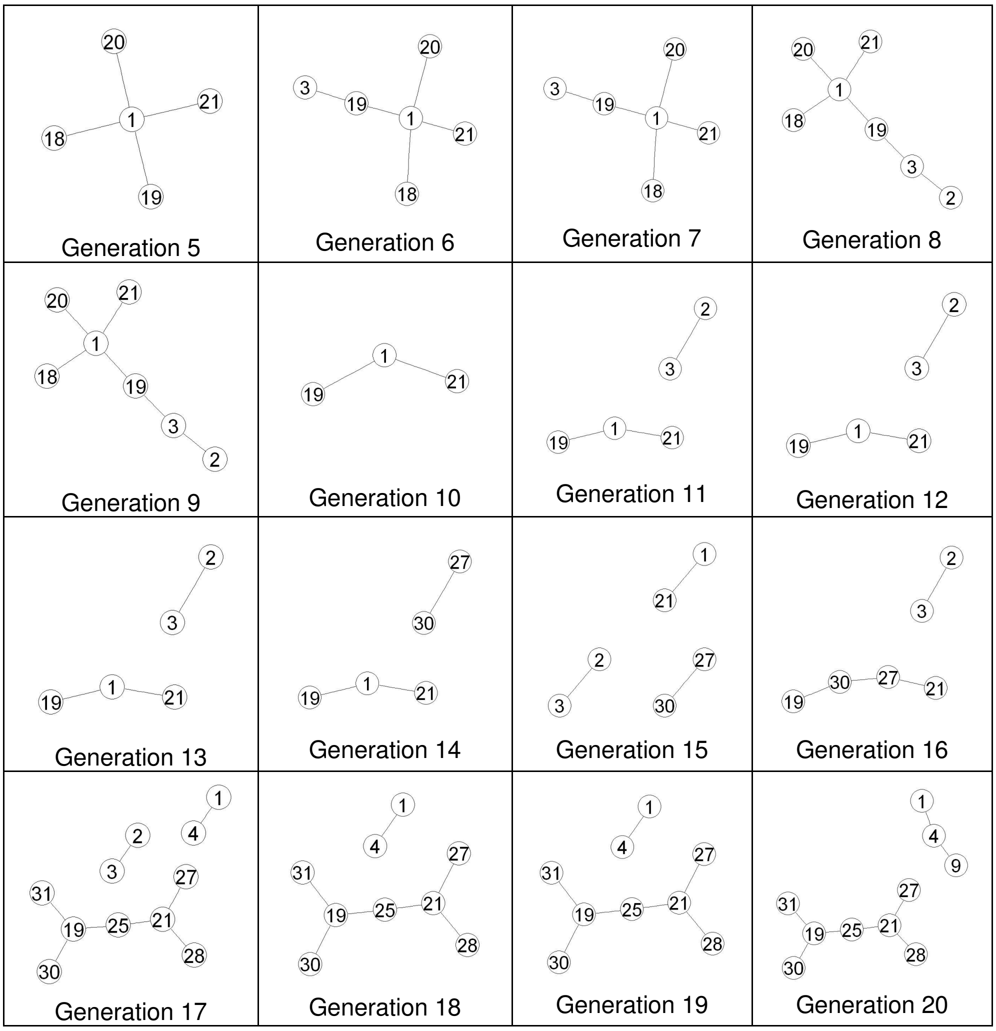


Figure 6-13 - Graphical evolution of the networks created via strategy f)

Figure 6-13 shows that the cooperative network that was formed in generation 5 joining firms 1, 18, 19, 20 and 21. The network changed its shape during the evolutionary process. In generations 6 and 8, firms 3 and 2, respectively, entered the network. The initial group of firms broke the contract in generation 10, but three of them keep the links between them (firms 1, 19 and 21). This may suggest that firms experiencing former collective networks can gain in terms of stock of knowledge that will be used in further relations with other (or with same) firms. In fact, firms 1, 19 and 20 of the initial group have links with others in almost all the generations until the end of the period of analysis.

*g) Virtual collaboration networks (***B2.1+a***)*

Strategies g) and h) demonstrate the situations in which the physical distance between the elements of the network has relatively low importance (only the technological distance is considered). Figure 6-14 illustrates the evolution of such *Virtual collaboration networks*.

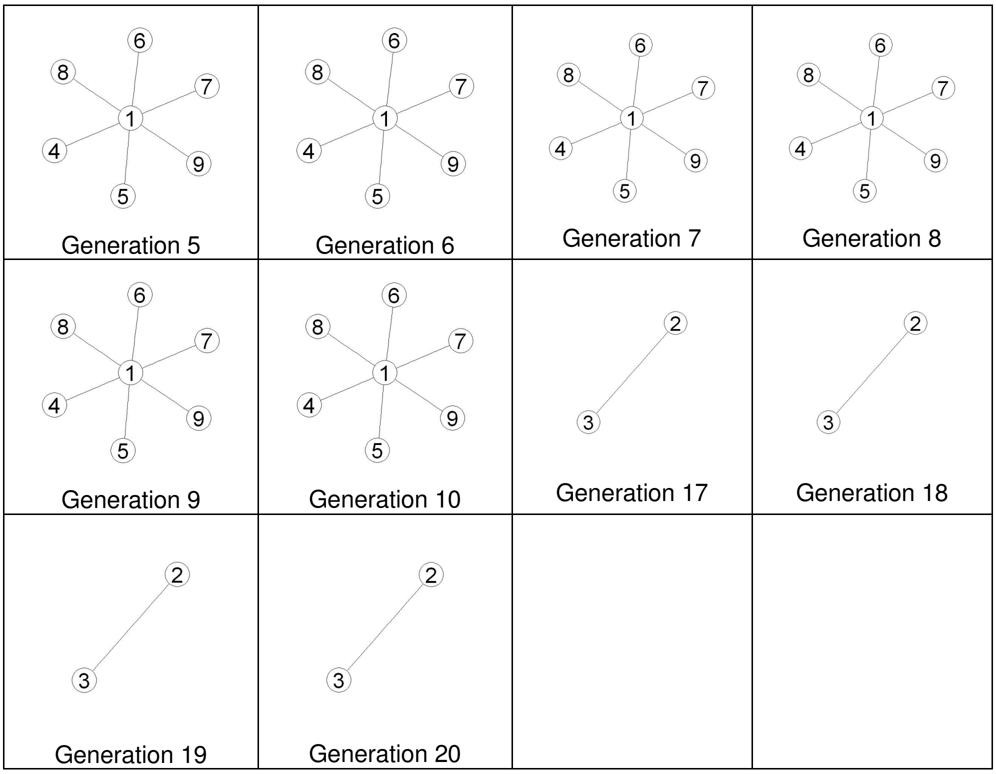


Figure 6-14 - Graphical evolution of the networks created via strategy g)

As in strategy e) (*collaboration networks*), we will assume that virtuall collaboration networks do not admit new elements throughout their existence. In the example of Figure 6-14, there is a vertical network including firms 1, 4, 5, 6, 7, 8, and 9 that was created in generation 5 and lasts until generation 10. There are no more networks until generation 17, in which a horizontal network (with firms 2 and 3) is created. This network lasts until the end of the period of analysis.

*h) Virtual cooperation networks (***B2.1+a***)*

As stated before, virtual cooperation networks are similar to cooperation networks, except that the physical distance has low importance here (only the technological distance is taken into account). For new elements that want to enter the network, they must negotiate directly with one of the elements of the network. In Figure 6-15 we represent the evolution of a virtual cooperation network.

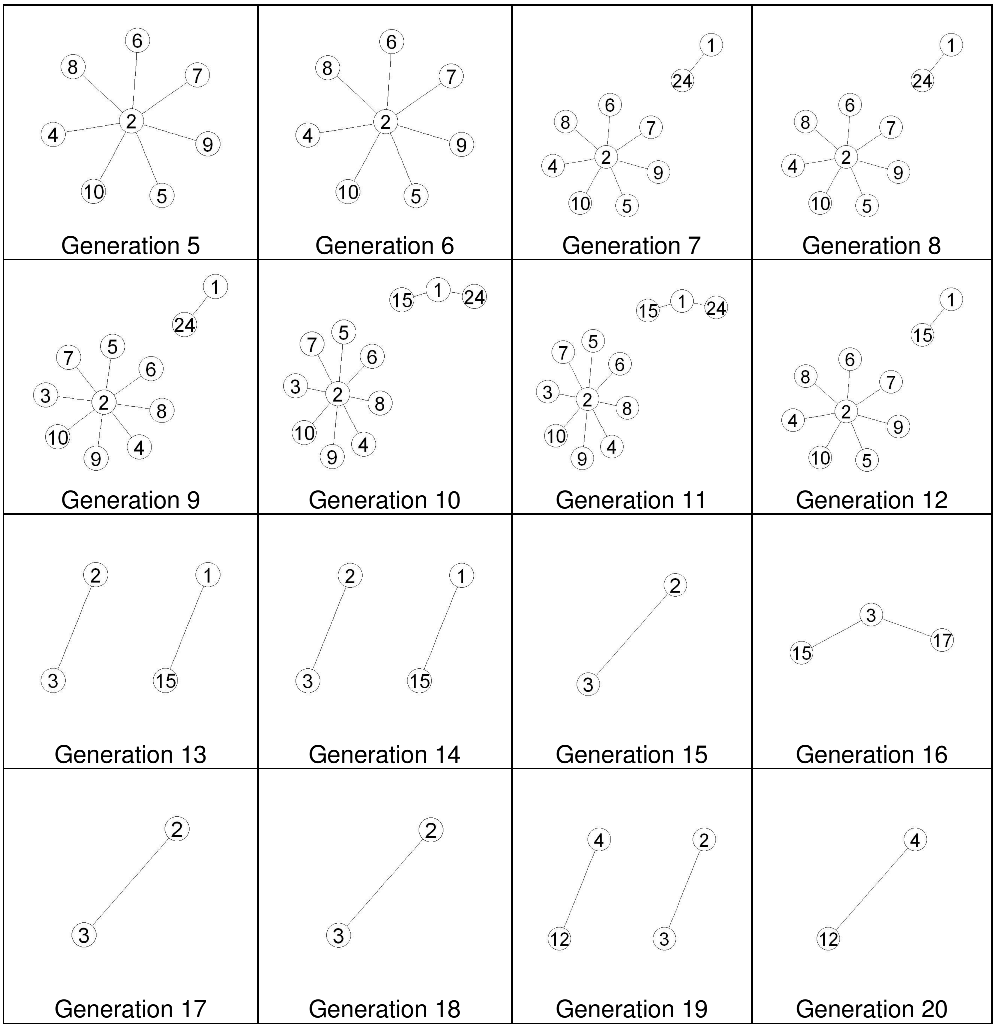


Figure 6-15 - Graphical evolution of the networks created via strategy h)

In the evolution depicted in Figure 6-15, a virtual collaboration network is created in generation 5 and broke at generation 13. Meanwhile, at generation 9, firm 3 entered the network by linking with firm 2 (the center of the network) but left the network at generation 12. Another network has been formed during this period: at generation 7, a network joining firms 1 and 24 (and firm 15, later on) was created. From generation 13, a new network arises (connecting firms 2 and 3) which is maybe a result of the gain in terms of stock of knowledge that was achieved previously by these firms.

Discussion

Table 6-9 shows some indicators associated with the simulations using different strategies. We have measured the profit, marginal cost, and the stock of capital of all of the networks that were created during the simulation runs[117].

Collaboration Strategies		Profit(*) (π^t)	Marginal Cost (c)	Stock of knowledge (**) (k^t)	Network structure indicators			
					Diameter	Path	Transitivity	Density (***)
a)	*Peer-to-peer complementariness*	3.27	0.0363	4247.8	1.19	1.07	0.00	2.22
b)	*Average stock scenario with great expected value*	4.61	0.0099	589.9	2.73	1.82	0.09	11.83
c)	*Preferential meeting process with Profit evaluation*	4.31	0.0155	1798.0	2.06	1.50	0.06	7.36
d)	*Concentration process*	4.55	0.0104	504.2	2.75	1.82	0.10	13.07
e)	*Collaboration networks*	2.52	0.0268	726.2	1.59	1.34	0.03	4.13
f)	*Cooperation networks*	3.70	0.0250	4322.7	1.67	1.35	0.01	4.22
g)	*Virtual collaboration networks*	2.68	0.0290	197.2	1.75	1.47	0.01	5.36
h)	*Virtual cooperation networks*	3.50	0.0264	2525.7	1.64	1.32	0.04	3.88

Table 6-9 –Comparison of collaboration strategies.

Notes: average values within each variable were computed for all variables within each strategy;

Only networks with three or more firms were considered in the analysis;
(*) Log(Profit); ; (**) Average stock of knowledge of three markets (X, Y_1 and Y_2); ; (***) Average number of nodes (firms) per network.

117 NetOrg was run ten times for each strategy many networks were produced. A file containing 4.113 networks produced to perform further statistical analyses. The average values of the networks within each strategy were taken for all the indicators. For the stock of capital we computed the average of the levels for the three industries, X, Y_1 Y_2.

The average of each variable (profit, marginal cost, stock of knowledge) in Table 6-9 was computed for each strategy as being the corresponding sum of the variable values divided by the number of networks within each strategy. The same average was applied to the network structure indicators such as diameter, average path length, transitivity and density[118] .

Strategies b) and d) seem to be the most profitable out of all. The average *profit* is greater than in other strategies, although the standard deviation is quite large to be able to come to a definite conclusion[119]. Profit is also high in strategy c). Marginal costs are lower, on average, in strategies b) and d), while strategies a) and e) reach the highest levels of marginal cost. Strategies a) (*Peer-to-peer complementariness*) and f) (*cooperation networks*) are the ones in which the stock of knowledge is higher.

Regarding network structure indicators, strategy a) has the lower diameter and average path length. Density (the average number of firms per network) is also very low when compared to other strategies (around 2 firms per network). The value of transitivity is zero in this strategy: a great number of the networks created under this strategy had two nodes only (and were not considered in the analysis) and the remaining networks within this strategy exhibit a linear form[120]. *Average stock scenario* (strategy b) and *Concentration process* (strategy d) present the highest values for all network structure indicators. These two strategies produce networks that are denser and more clustered, with higher diameter and average path length.

[118] As presented above in Chapter 2, network statistics are important to quantify the structure of the networks. The *diameter* of a network is a measure of the network length: diameter is the distance, in number of nodes, between the two farthest nodes of the network; the *average path length* or connectivity length (or just *path*) is the mean number of steps that are necessary to move in the shortest route between two nodes; *transitivity* is a network property that measures the connectivity (or clustering) inside the network. Finally we used the number of nodes in each network as a measure of *density***.** A different measure could bee used here (as the number of links divided by the number of nodes), but we decided to choose the simplest way to compute the network density.

[119] In what concerns the *profit*, the coefficient of variation, σ_i/μ_i, where σ_i is the standard deviation and μ_i is the mean of the population of all networks under any strategy i, stands between 0.27 (strategy d) and 0.95 (strategy e). The variability of the *stock of knowledge* is higher (mainly in market X) since the networks within the same strategy present different patterns of evolution.

[120] Transitivity measures the clustering or connectivity inside the network. As expected, Linear networks show lower values of transitivity, in average, than Star and Multipolar networks, as showed in the following table:

	Form		
	Linear	Star	Multipolar
Transitivity	0,0085	0,0414	0,2577

There is, in fact, a positive relationship between network structure indicators. The association between them is represented in Table 6-10:

	diameter	path	transitivity	density
diameter	1			
path	0.799	1		
transitivity	0.449	0.316	1	
density	0.598	0.674	0.392	1

Table 6-10 - Pearson correlation coefficients between network structure indicators

It can be observed that strategy a) generates low profit, when compared to the others, but the corresponding stock of knowledge is one of the greatest among all strategies. We recall that, according to Equation 6-16, there is a cost associated with the creation of knowledge, which may decrease the profit value in certain cases such as in this strategy. However, in general, there is a significant positive correlation between profit and stock of knowledge - Pearson correlation coefficient is 0.495(6).

In order to extract patterns from the data produced by network evolution, we have performed a classification with decision trees, using C5.0 (Quinlan, 1997). Emerging patterns can be used to gain insight into aspects of network collaboration. We used the same data file (used before to compute network indicators presented back in Table 6-9), containing 4.113 networks, produced by the eight strategies. In the first case, the eight strategies were analysed all together. The same process of rules extraction was done separately for each strategy afterwards.

All the variables (profit, marginal cost, stock of knowledge - split for markets X, Y_1 and Y_2 -, diameter, path, transitivity and density) have been categorized into three different levels, according to their original values: 1-low; 2-medium; 3-high. A method of equal frequency categorization has been used. Profit was used as the classification (or output) variable. The designations for the new categorized variables are described in the following table:

Original name	Name of categorized variable
Profit	Class_Profit
Marginal Cost	Class_MargCost
Stock of Knowledge (market X)	Class_StockX
Stock of Knowledge (market Y_1)	Class_StockY_1
Stock of Knowledge (market Y_2)	Class_StockY_2
Diameter	Class_Diam
Path	Class_Path
Transitivity	Class-Transit
Density	Class_Density

Table 6-11 - Designations for the categorized variables used in the classification problem with C5

Two of the main extracted rules (out of 21 rules, with a minimum accuracy of 87.8%) are summarized in the following lines (confidence levels are presented in brackets):

```
IF Class_Transit = 1
     THEN Class_Profit=1  [0.935]

IF Class_StockY1=3 & Class_StockY2=3 & Class_Diam=2
     THEN Class_Profit=3 [0.878]
```

These rules indicate that less profitable networks have lower transitivity (small clustering coefficient). Simulaneously, more profitable networks are associated with greater stock of knowledge and intermediate diameter.

We have also produced rules for each strategy separately in order to look for specific patterns that are not found in the overall analysis. Table 6-12 presents the most important rules for each of the eight strategies. Rules are organized according to the strategy, attributes and the output classification value is in column *Class_Profit*:

Strategy	Class-StockX	Class_StockY$_1$	Class_StockY$_2$	Class_Marg Cost	Class_Diam	Class_Trans	Class_Path	Class_Profit	Confidence
a)						1		**1**	0,947
a)						2		**2**	0.947
a)						3		**3**	0.947
b)		2		2	1	2		**3**	0.857
b)				1		2		**3**	0.837
b)		3						**3**	0.706
c)	1				2			**1**	0.981
c)	{2,3}				2		{1,2}	**3**	0,991
c)					2	1		**1**	0.978
c)		1		3		1		**1**	0.976
c)					2		3	**1**	0.969
d)	{2,3}	1				1		**2**	0,9
d)					3		2	**3**	0.697
e)	3					{1,2}	{2,3}	**2**	0.95
e)					1			**1**	0.992
e)					1			**1**	0.992
e)					{2,3}	3		**3**	0.991
e)					3		1	**3**	0.98
f)						1		**1**	0.963
f)					{2,3}	3		**3**	0.995
g)					1			**1**	0.957
g)					{2,3}	{1,2}		**2**	0.917
g)					{2,3}	3		**3**	0.95
h)					{2,3}	3		**3**	0.895
h)						1		**1**	0,951

Table 6-12 – Summary of the main rules produced by C5 for the classification of Profit

Note: all variables were categorized assuming the values: 1-low; 2-medium; 3-high

Table 6-12 shows the rules with higher confidence values chosen among all the rules produced within each strategy. Profit has been classified in classes 1 (low) to 3 (high). We can observe that in general, when profit is high the stock of knowledge is high too, and in some corresponding cases, the marginal cost is low. On the other hand, we can observe that, in general, when profit is low the stock of knowledge is low too, and in some corresponding cases the marginal cost is high.

In relation to other variables, there is not a clear emerging relationship: for example, rules shown in Table 6-12 seem to indicate that profit is inversely related to the average path length and, positively related to transitivity and diameter. In fact, most part of the rules that are associated with high profit has high values in transitivity.

We were also concerned with the evolution of the network indicators along the generations. Figure 6-16 illustrates the evolution of the profit across the networks lifetime.

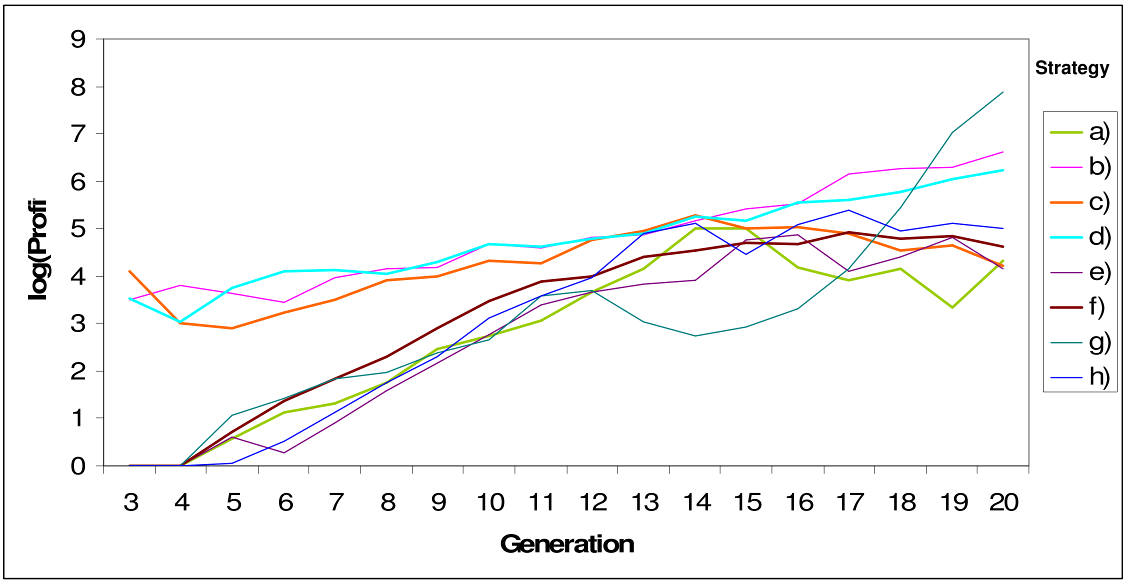


Figure 6-16 - Evolution of network profit (logarithm of profit) in all strategies along the twenty generations of the simulation

(Note: the average profit of all networks within the same strategy was considered along the generations)

We can observe that the profit increases, in average, along the generations. As we have seen above, links between the firms enhance the diffusion of the knowledge within the network and the profit increases as well. It is interesting to observe that some strategies, as g), which seem less profitable in the first generations, increase drastically their profit at the end of the period.

In Figure 6-17 we depicted the evolution of networks' average diameter over time for all the strategies. In all strategies we can observe that the diameter reduces along the generations.

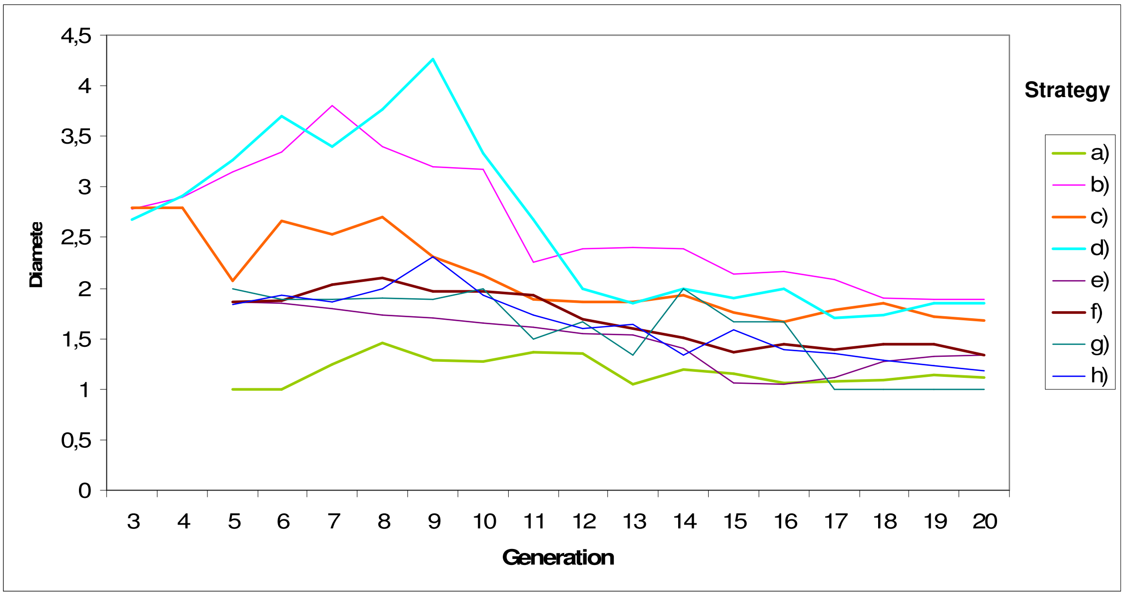


Figure 6-17 – Evolution of network diameter in all strategies along the twenty generations of the simulation

(Note: the average diameter of all networks within the same strategy was considered along the generations)

Strategy d) (*concentration process*) creates larger networks, in average, as we have seen before, although the diameter decreases along the iterations. The networks created with the *peer-to-peer complementariness* (strategy a), which we have seen that generate higher stock of knowledge and have lower diameter, slightly reduce their diameter, in average, during their generations.

The increase of the profit and the corresponding reduction of the diameter all along the generations, is in line with what is said in literature about economic networks, i.e., that the diameter of networks decreases along time. (Leskovec *et al.* 2005). This is actually one property of the small-world networks, as we have seen in Chapter 2.

Concerning the evolution of transitivity, it is important to note that it tends to increase strongly in the initial life of networks and tends to decrease slowly in the end of the twenty generations. This latter phenomenon is a consequence of the reduction on the network dimension that is observed in the final of the period of analysis.

In the following we will explore the outcome of other effects (introduced as user defined parameters) such as migration, attitude to risk, network support and learning on the characteristics of the networks.

Location and Migration

There are some cases where firms connect with other firms in other regions. In the next example, some of the networks were created involving firms from both regions. Figure 6-18 shows four networks that involved firms from both regions:

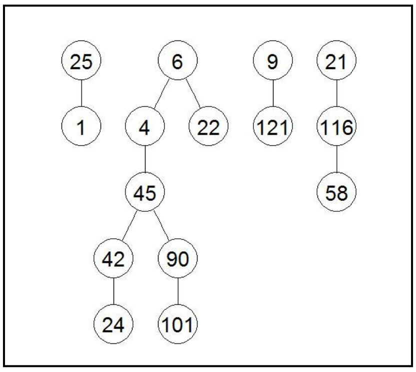


Figure 6-18 – Networks with firms from different regions

Firms migrate in order to increase their profit. We observed that firms and networks proliferate in region 2, the region with lower marginal costs. Some of the firms that migrated (from region 1 to region 2) were able to overcome negative profits. Some of them have increased their profits considerably and became top success firms, within higher profit and stock of knowledge. Although it seems that firms with higher values of risk are in general more profitable, it appears that there is no direct association between the profit of the firm and its level of risk (Pearson correlation coefficient revealed a weak value of -0.067).

Figure 6-19 shows networks in the place where they have been created or in the new place after firm dislocation. We can observe that there are more networks in region 2 than in region 1, as a result of the lower marginal costs in that region. Nevertheless, there are some firms and networks in region 1, as the firms' clustering helps avoiding negative results.

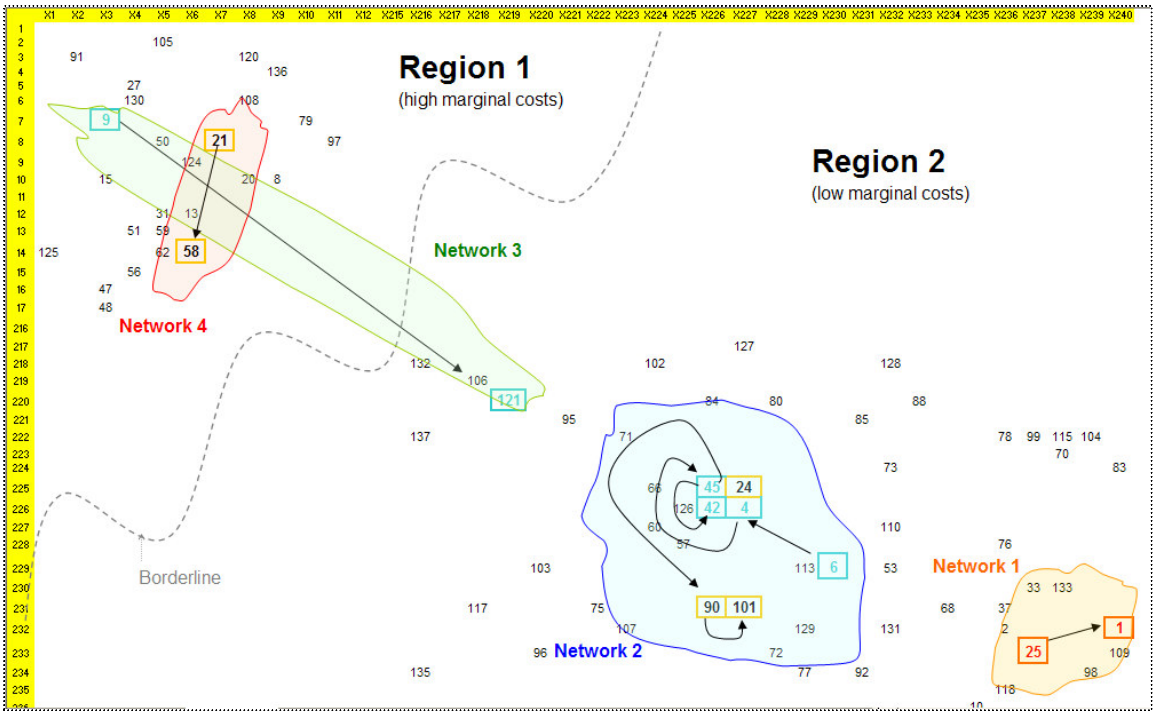


Figure 6-19–Representation of the space showing the firms (in numbers), the links (arrows) and the networks (in colored areas) in the 15th generation

Note: the direction of the edges in the arrows are arbitrary and do not indicate the direction of the links.

Attitude to risk

In the simulation we consider that firms have different attidudes to risk. As we have seen before, the attitude to risk is associated with migration and with saving partner firms from failure, that is:

- more risky firms tend to migrate to a different region when they are in a difficult situation, and
- more risky firms tend to save other firms from failure, even if it brings costs to them.

As said in Section 6.4.6 two kinds of attitude were considered: "*conservative*" and "*adventurous*". We use the term "*risk level*" to represent this attitude. A certain level of risk was therefore defined for every firm (we considered an integer value from the interval [1, 10]: *conservative* firms have risk levels under 5, while *adventurous* have these levels greater than or equal to 5. A risk level is assigned at random to each firm and it is not changed during its lifetime. Table 6-13 summarizes the average profit obtained by firms, according to the levels of risk within the eight collaboration strategies:

Collaboration strategies	Profit* average Conservative firms	Adventurous firms
a)	1.4	1.6
b)	1.9	3.2
c)	1.0	4.2
d)	1.9	4.0
e)	2.6	2.7
f)	0.7	3.1
g)	0.8	2.6
h)	0.4	3.4

Table 6-13 – Average of firms' profit according to the *attitude to risk*

Notes: All firms have been considered independently of being located in a network or not;

(*) Logarithm of Profit

These results are not directly comparable with those of Table 6-9, because those were computed only for the firms in the networks, while the former were computed for all firms of the simulation. We observe that adventurous firms seem to perform better than conservative ones. The Mann-Whitney Test (Conover, 1999) was used to confirm that profit is lower in conservative firms (p-value<1%)[121]. This is really to be expected as it is a consequence of the built-in strategy (adventurous firms should migrate to regions with low marginal costs). The effects of migration to regions with low marginal costs for adventurous firms are clear and make them increase their profits.

[121] The Mann-Whitney test (Conover, 1999), also known as the Mann Whitney non parametric U test involves studying the position or ranking of data from two independent samples (defined by 1 and 2). The null hypothesis (H_o) for this type of test is: $\mu_1=\mu_2$, which states that there is no difference between the samples 1 and 2, μ_1 and μ_2, representing respectively the medians (or means in the case where the samples are both symmetric) of the samples 1 and 2. Usually, the alternative hypothesis H_1 is bilateral ($\mu_1 \neq \mu_2$) meaning that there is a difference the two samples. However, other alternative hypotheses can be defined ($\mu_1<\mu_2$ or $\mu_1>\mu_2$) depending on the types of analyses. The conclusions can be taken in terms of the results of p-values, in which a p-value lower than 5% indicates that the null hypothesis shall be rejected, meaning that the difference between the two populations is significant. When the p-value is lower than 1% we shall conclude that those differerences are highly significant. This is a non parametrical alternative to the t-Test for the comparason of means (that is used under the normal assumption).

Network support

Network support is a mechanism that also has been introduced into NetOrg. In some types of network, essentially in those where a common goal exists, collective efforts are exerted in order to save firms from failing. This is one of the firms' advantages of being *linked,* since relationships are established between firms existing in the same network[122].

The way how we have implemented *network support* in NetOrg is simple: if this mechanism (as well as birth and death parameters) is active, and if any firm, say, **i**, is about to fail due to the predefined conditions (see Section 6.4.1), then efforts are made to help the failing firm. These efforts are made by the firms that are connected to **i** and whose attitude to risk is positive (risk level greater than 5). The action consists in transferring some of their knowledge, materialized under the form of stock of knowledge to the potentially failing firm.

Network Support was not considered active in any of the scenarios associated with the previous simulation outputs. To analyse the impact of the *network support* function we run the same series of experiments maintaining the previous set of parameters, except for *network Support*, that has been activated in this case. All the collaboration strategies have been used. Table 6-14 shows the results for networks concerning both situations:

[122] As seen in Section 2.3., based on Hakansson approach, Ratti (1991) states that the cohesion of the network is determined by certain specific forces that are very important to the survival of the network and can be understood as the mechanism of the relationships that rules the network.

Network Support			
Strategy	Profit* (π^t)	Marginal Cost (c)	Stock of Knowldege (k^t)
a)	5.3	0.0084	846.4
b)	4.4	0.0099	355.4
c)	4.4	0.0072	277.3
d)	4.6	0.0087	342.2
e)	2.6	0.0094	240.2
f)	3.4	0.0157	210.8
g)	1.4	0.0244	118.4
h)	2.4	0.0341	222.8
average	*3.6*	*0.0147*	*326.7*
No Network Support			
Strategy	**Profit* (π^t)**	**Marginal Cost (c)**	**Stock of Knowldege (k^t)**
a)	3.3	0.0363	4247.8
b)	4.6	0.0099	589.9
c)	4.3	0.0155	1798.0
d)	4.5	0.0104	504.2
e)	2.5	0.0268	726.2
f)	3.7	0.0250	4322.7
g)	2.7	0.0290	197.2
h)	3.5	0.0264	2525.7
average	*3.6*	*0.0224*	*1864.0*

Table 6-14 – Networks' results according to *network support*

Note: The average of the three markets (X, Y_1 and Y_2) was considered for the stock of knowledge
(*) Logarithm of Profit

Marginal costs are higher, on average, when the *network support* is not activated. The average stock of knowledge is also higher when the *network support* is not activated. To compare the results of networks, we have performed a Mann-Whitney non parametric test, and concluded that the difference between the two situations (*Network Support* vs *No Network Support*, taken as the grouping variable) is significant for the Marginal Cost (p-value<1%) and Stock of Knowledge (p<1%) but it is not significant concerning the Profit (p-value>5%).

Learning

Learning is related to the question G5, formulated in Chapter 4: "*What is the effect of individual learning in the formation of networks*?". The learning mechanism implemented in NetOrg is based on instance-based learning. We recall that the learning mechanism allows firms to evaluate the performance of the potential partner and to decide whether it would be a good link to consider before taking the final decision. A firm i decides to establish a link with a new partner (say, firm j), based on previous information about the success of that partner (or other similar partners) already experienced by other firms. This information includes the level of the stock of knowledge, K^t_j, and its profit, π^t_j.

To measure the effects of learning in the firms of NetOrg we have compared the indicators for the following two situations: when firms learn and when they do not. Table 6-15 present a summary of the results of the networks according to *learning*.

Learning			
Strategy	**Profit* (π^t)**	**Marginal Cost (c)**	**Stock of Knowldege (k^t)**
a)	3.3	0.0363	4247.8
b)	4.6	0.0099	589.9
c)	4.3	0.0155	1798.0
d)	4.5	0.0104	504.2
e)	2.5	0.0268	726.2
f)	3.7	0.0250	4322.7
g)	2.7	0.0290	197.2
h)	3.5	0.0264	2525.7
average	*3.6*	*0.0224*	*1864.0*
No learning			
Strategy	**Profit* (π^t)**	**Marginal Cost (c)**	**Stock of Knowldege (k^t)**
a)	3.8	0.0290	275.4
b)	5.3	0.0096	6028.9
c)	5.0	0.0074	725.2
d)	5.1	0.0064	38254.7
e)	2.4	0.0400	304.8
f)	4.1	0.0147	2754.1
g)	3.4	0.0161	3084.6
h)	2.9	0.0149	691.0
average	*4.0*	*0.0172*	*6514.8*

Table 6-15 - Networks' results according to *learning*

Note: The average of the three markets (X, Y_1 and Y_2) was considered for the stock of knowledge
(*) Logarithm of Profit

Learning does not increase the profit of the firms in the networks, as well as their stock of knowledge. This background knowledge seems not to be of great importance for the decision making in the linkage process. In fact, there is statistical evidence that learning decreases the values of the profit (p-value<1%, in the Mann-Whitney independency test) and increases the marginal cost (p-value<1%). On the other hand, the difference in the values of stock of knowledge is not statistical significant (p>10%).

However, results differ, if we analyse the population of firms instead of networks. Considering all the firms in the simulation that are linked, learning slightly increases their profit although this increase is not significant (p-value>86%). The difference between the stock of knowledge (for Market X) in the two situations (learning and no learning) is significant (p-value<5%), with the stock being higher in the learning situation. We can therefore conclude that the learning mechanism improves the stock of knowledge if we consider the linked firms, but this is not true if we take networks instead of firms[123].

In the following sections, we proceed with the validation of the model, using statistical tests and direct observation from results.

6.9.3. Validation of the model: some hypotheses testing

Previously we saw that different values of built-in parameters effectively lead to results that are quantitatively and qualitatively different. After defining the initial values for some parameers and obtaining the results of the simulation, it is now necessary to test the robustness of the hypotheses, as given in Section 3.3.3 and Section 6.8.

Therefore, to verify that the model is able to reproduce some aspects of reality associated with the target industry, we are going to use the facts used in Section 6.8. (Identification of the hypothesis). In the following sections we will review these hypothesis and perform their testing. Hypotheses have been grouped according to different types of analysis. The first set of hypotheses, H_{21} and H_{23}, are confirmed by the empirical evidence of data, while hypotheses H_{22}, H_{29} and H_{28} will be tested using the Mann-Whitney non-parametrical tests (that is used to compare independent samples). Finally, we use survival analysis for firms and after that for networks, in order to test hypotheses H_{24}, H_{25}, H_{26} and H_{27}.

[123] Although learning has been modelled as an individual process at the firm level, its impact can be measured either at firm or network level.

In all these cases we have maintained the same parameters: Neighbourhood *Ring*=3, the set of density interval limits, and activated *Learning*, *Network Support* and *Migration,* (some of the user defined parameters). All the collaboration strategies have been considered.

Concentration and Migration

Starting with empirical observations H_{21} and H_{23}, we recall that:

> **H_{21}:** *Buyers promote both a concentration process of suppliers (…)*;
>
> **H_{23}**: *Migration of firms to markets with lower marginal costs is a way to rationalize production.*

Facts H_{21} and H_{23} are easy to verify and thus do not need any statistical confirmation. In fact, in Section 6.9.3 we saw that the simulation has produced several networks; some of them are concentrated in one ore more industries of type X, mainly when the strategy d) (*concentration process*) is used.

When the migration option is active, some firms also opt to dislocate to markets with lower marginal costs. We have observed that some of the adventurous firms[124] that migrate from region 1 to region 2, were able to overcome their situation of negative profits. Some of them increased considerably their profit and became top success firms.

Comparing indicators for firms in and out of the networks

The next hypothesis that we considered is H_{22}. We have verified its validity by comparing performance indicators for firms that are included in networks with firms that are not linked to any network.

> **H_{22}:** *Some collaborative firms have increased their profits because of the reduction of costs due to the rationalization of some of their functions [through collaboration via network formation];*

[124] These are the firms with positive attitude to risk.

In Table 6-16 we compare the indicators for firms linked and not linked in networks. The performance of linked firms is better than not linked ones. The former have higher profit and stock of knowledge, and lower marginal cost, in average.

Firms in networks			
Strategy	Profit* (π^t)	Marginal Cost (c)	Stock of Knowldege (k^t)
a)	5.8	0.0242	16409.6
b)	6.0	0.0157	6935.7
c)	7.9	0.0146	58539.7
d)	7.9	0.0086	21023.4
e)	5.4	0.0351	170640.6
f)	8.2	0.0191	43717.6
g)	2.8	0.0383	994.3
h)	5.6	0.0141	75827.3
average	*6.2*	*0.0212*	*49261.0*
Firms not in networks			
Strategy	Profit* (π^t)	Marginal Cost (c)	Stock of Knowldege (k^t)
a)	0.9	0.0193	42.7
b)	-0.3	0.0498	3.9
c)	-0.4	0.0536	1276.0
d)	-0.2	0.0507	1110.5
e)	1.1	0.0218	19.4
f)	0.0	0.0315	4.4
g)	1.3	0.0295	179.6
h)	0.0	0.0206	4.9
average	*0.3*	*0.0346*	*330.2*

Table 6-16 - Networks' results obtained for firms linked in networks and not linked

Note: The average of the three markets (X, Y_1 and Y_2) was considered for the stock of knowledge

(*) Logarithm of Profit

Table 6-16 shows that the performance of the firms belonging to networks is different from the performance of firms not connected to networks. We concluded that the difference between these indicators in the two situations (*Firms in networks* and *Firms not in* networks) is significant for all the variables[125]. As expected, marginal costs are reduced and the stock of knowledge is greater for firms in networks. Besides, this was indeed expected, as networks should help to spread innovation and therefore increase the profit.

[125] The Mann-Whitney test of independence presented the following p-values: Profit (<0.1%), marginal cost (5%), stock of knowledge (0.1%)

To compare the life expectancy in the two situations, we have activated the user defined parameters *birth* and *death*. Then, we observed that the average age of firms in networks is higher (13.7 periods) compared to the not linked firms (8.7 periods)[126].

Now, we will test the observations associating the type of cooperation (horizontal/vertical[127]) with the distance between firms:

> **H_{29}:** *If the distance between two firms is short, they will have a higher tendency to cooperate horizontally; if the distance between two firms is high, they will have a higher tendency to cooperate vertically.*

In what concerns the hypothesis H_{29}, we have interpreted it as a comparison of the distances (technological, geographical and total) between firms, taking into account the type of cooperation – horizontal and vertical. We recall that cooperation can be horizontal if it involves firms with the same type of product/market, while vertical cooperation occurs when it involves firms with different types of products /markets. Since the samples of the distances do not follow the normal distribution, a non-parametrical Mann-Whitney test has been carried out to test the following set of hypotheses:

H_o: $\mu_1=\mu_2$

H_1: $\mu_1<\mu_2$

where μ_1 represents the mean[128] distance between firms having *horizontal cooperation,* while μ_2 represents the mean distance between firms having *vertical cooperation*. So, if we reject the null hypothesis, H_o, this indicates that the mean distance between firms with horizontal cooperation is lower than that of firms with vertical cooperation. A grouping variable, *type,* was defined in the following way: type=1 (horizontal cooperation); type=2 (vertical cooperation). The following tables show the results of

[126] We used the *average age* as a proxy of *life expectancy.*

[127] As we said before, we assume that horizontal links are connections between firms within the same market, while vertical links are connections between firms from other markets.

[128] In several non parametrical statistical tests such as Mann-Whitney, the median is the parameter to be tested. However, since we assume that there is symmetry in data, the median and the mean are considered equivalent.

this test, showing the three types of distance: Geographical, Technological, and Total obtained from a set of 151 networks:

(a) Ranks

	Type	N	Mean Rank	Sum of Ranks
Geographical distance	1	90	75.21	6768.50
	2	61	77.17	4707.50
	Total	151		
Technological distance	1	90	70.09	6308.00
	2	61	84.72	5168.00
	Total	151		
Total distance	1	90	71.47	6432.00
	2	61	82.69	5044.00
	Total	151		

(b) Test Statistics

	Dgeo	Dtec	Dtotal
Mann-Whitney U	2673.500	2213.000	2337.000
Z	-.333	-2.017	-1.547
Exact Sig. (1-tailed)	.372	.022	.061

Table 6-17 - Ranks (a) and test statistics (b) of the Mann-Whitney Test for the comparison of distance between networks according to the type of relationship (1 - horizontal vs. 2 - vertical)

We can observe that when variable *type* takes the value 1 (horizontal cooperation), the mean rank of the geographical distances is lower in all distances (geographical, technological, and total) than when *type* takes the value 2 (vertical cooperation). However, only the difference concerning technological distance is significant (p-value<5%). This means that for technological distance, the evidence H_{29} was confirmed: if the distance between two firms is short, they will have a higher tendency to cooperate horizontally; if the distance between two firms is high, they will have a higher tendency to cooperate vertically.

Finally we test hypothesis H_{28}:

> **H_{28}:** *Firms that have long-term relationships with other firms will live longer, in average*

To test this hypothesis, we activated the *birth* and *death* user defined parameters and collected information of firms during the runs concerning all strategies. We registered the following data: *Age* of firm, to measure the duration of life, *Duration* of link (both measured in time periods), and Firm *Status* (dead=0 or alive=1). Table 6-18 presents a short statistical summary of the variable *duration of link.*

	N	Minimum	Maximum	Mean	Std. Deviation
Duration of link	78	3	131	10.60	17.368

Table 6-18 – Descriptive statistics of variable *Duration of links*

The maximum value of the duration of link should be, apparently equal to the number of generations, that is, 20. However, the numbers we present are the cumulative durations of all the links connected to a particular firm.

In order to compare the survival of the firms in both situations of long-term and short-term relationships, we created a new variable (*relationship*) that discriminates these two kinds of link duration. If the total duration of all the links of a particular firm is greater than 10 periods, we consider it a long-term relationship otherwise it is a short-term relationship. This threshold is the value of the mean of *relationship*.

Table 6-19 shows some statistical indicators of the age of firms, taking into account these two types of relationships:

Relationship	Age of firms		
	Mean	Std. Error	Median
Short-term	7.01	0.563	6
Long-term	11.80	2.407	16

Table 6-19 – Descriptive statistics of Age according to the type of relationship

It seems that firms that have long-term relationships live longer, in average. To confirm this hypothesis, a non-parametrical Mann-Whitney test has been performed again to test the following set of hypothesis:

H_o: $\mu_1=\mu_2$

H_1: $\mu_1<\mu_2$

where μ_1 represents the mean duration of the links for firms with *long-term relationships* and μ_2 represents the mean duration of the links for firms with *short-term relationships.* The *duration of the links* is measured by the *age* of firms. The null hypothesis was rejected (p<5%).

Ranks (a)

Relatioship	N	Mean Rank	Sum of Ranks
Short-term	68	37.59	2556.00
Long-term	10	52.50	525.00
Total	78		

Test Statistics(b)

	Age
Mann-Whitney U	210.000
Z	-1.966
Exact. Sig. (1-tailed)	.025

Table 6-20 - Ranks (a) and test statistics (b) of Mann-Whitney Test for for the comparison of the duration of the links among firms with long-term and short-term relationships.

From the simulated data it can be concluded that the difference in the duration of the links for firms with long-term and short or medium-term relationships is significant. Therefore, we can say that firms having long-term relationships liver longer in average.

We also performed a survival analysis using the original variable *duration of link* as a covariate, but the result was not significant (p-value>70% in the omnibus test of model coefficients).

6.9.4. Validation with Survival Analysis

As introduced in Section 5.6.2, Survival Analysis is a way to assess the impact of the variables on the survival of individuals. In our case, we will use Survival Analysis to explore the impact of some independent variables (here called covariates) when considering the mortality of firms and networks. We split our study in two parts according to the object of analysis: first, we will analyse the impact of some variables on the survival of networks. Next, we study the impact of some variables on the survival of individual organizations.

Survival of Networks

In order to explore the impact of some independent variables in the mortality of networks[129], we have collected 501 networks[130], each of which being characterized by a set of 13 attributes that describe the features of a *particular network* in the problem:

- Profit (*profit*)
- Marginal Cost (*mcost*)
- Stock of knowledge in Market X (*StockX*)
- Stock of knowledge in Market Y_1 ($StockY_1$)
- Stock of knowledge in Market Y_2 ($StockY_2$)
- Age (*Age*)
- Form of the network (*Form*=1,2)
- Number of existing networks at the time of its birth (*netbirths*)
- Number of existing networks at the time of its death (*netdeaths*)
- Number of existing nodes (firms) at the time of its birth (*BNodes*)
- Number of existing nodes (firms) at the time of its death (*DNodes*)
- *Status* (dead=0 or alive=1)
- Collaboration Strategy (*Strategy*=a), b),…, h))

Except for the two nominal variables *Form*, and *Strategy*[131] and for *Status* (which is the state variable – with categories: death or alive - that determines the censored cases), all the rest are continuous or discrete attributes that can be treated as scale variables.

[129] The failure of a network occurs when it is not possible to continue following its evolution. As stated before, some networks are created, merged or deleted, during the several generations of NetOrg. For example, when two networks merge into one only, then we consider that the resulting network maintains the identification of the network that contains the most part of the nodes (between the two merging networks). In this situation the other network disappears.

[130] To collect these observations of networks, we have run NetOrg once for each different strategy, taking, for each strategy, 10 runs containing 20 generations each.

[131] In survival analysis, independent variables (covariates) can be continuous or categorical; if categorical, they must be dummy or indicator-coded. The impact of the latter in the Hazard function is measured by a contrast method, in which the different categories are compared with a reference category. In our case, *Form* and *Strategy* are indicator-coded categorical variables, and one new category, the last (the reference one) was created for each variable: category 4 in *Form* and category i) in *Strategy.* These reference categories contain the average of the remaining categories. Then, the *deviation* contrast method has been considered, in which each category of the predictor variable except the reference category is compared to the overall effect.

Variable *Form* indicates the shape of the network: it takes the value 1 when the shape is *linear,* the value 2 when it is a *single star*[132] *and* the value 3 when it is a *multipolar star*.

Covariates *netbirths* and *netdeaths* are collected in order to test, respectively, hypotheses H_{26} and H_{27}, considering the contemporaneous density and founding density:

> *H_{26}: Contemporaneous density (of networks) has a negative impact on the mortality of the networks*
> *H_{27}: Density (of networks) at founding has a positive impact on the mortality of the networks*

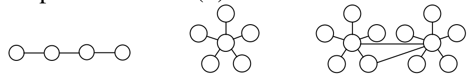
We also aim at testing hypotheses H_{210} and H_{211}:

> *H_{210}: Contemporaneous density (of firms) has a negative impact on the mortality of networks*
> *H_{211}: Density at founding (of firms) has a positive impact on the mortality of networks*

Just before employing the Cox Regression (we used SPSS version 15.0 again to perform survival analysis), we have tested the Proportional Hazard (PH) assumption. Since the Partial residuals are not normally distributed, although they are quite symmetrically distributed, we have computed the Spearman correlation between the Rank of Age and each one of the covariates' partial residuals. We recall that this step aims at investigating if the hazard rate is constant over time, and therefore, the *Age* is used as a proxy of the *time*.

Those networks that still existed at the end of the simulation have been censored from the database and were not considered in the analysis. Therefore all firms that were still

[132] Representation of the simplified forms of networks, according to Chapter 2: line (1), star (2) and multipolar network (3):



Form (1) – Linear network   Form (2) – Star network   Form (3) – Multipolar network alive in the moment of the final observation (i.e., the 20th generation) were censored. Table 6-21 presents the Spearman correlations between the *Schoenfeld partial residuals* (obtained for every covariate) and the ranked order of survival time (Age) for not-censored observations.

		Rank of *age*	Partial residual for *profit*	Partial residual for *mcost*	Partial residual for *stockX*	Partial residual for *stockY*$_1$	Partial residual for *stockY*$_2$	Partial residual for *netbirths*	Partial residual for *netdeaths*	Partial residual for *BNodes*	Partial residual for *DNodes*
Rank of *age*	Correlation Coefficient	1.000	-.034	.108(*)	-.160(**)	-.023	-.189(**)	-.163(**)	-.089	.009	-.144(**)
	Sig. (2-tailed)	.	.510	.035	.002	.652	.000	.001	.082	.866	.005
Partial residual for *profit*	Correlation Coefficient	-.034	1.000	-.331(**)	.452(**)	.128(*)	-.073	-.143(**)	-.111(*)	-.094	-.176(**)
	Sig. (2-tailed)	.510	.	.000	.000	.013	.155	.005	.031	.068	.001
Partial residual for *mcost*	Correlation Coefficient	.108(*)	-.331(**)	1.000	-.100	-.051	-.096	-.054	-.092	-.063	-.081
	Sig. (2-tailed)	.035	.000	.	.053	.327	.063	.296	.074	.221	.116
Partial residual for *stockX*	Correlation Coefficient	-.160(**)	.452(**)	-.100	1.000	.259(**)	-.144(**)	-.122(*)	-.138(**)	-.059	-.041
	Sig. (2-tailed)	.002	.000	.053	.	.000	.005	.018	.007	.253	.422
Partial residual for *stockY*$_1$	Correlation Coefficient	-.023	.128(*)	-.051	.259(**)	1.000	-.008	.017	-.137(**)	.030	-.151(**)
	Sig. (2-tailed)	.652	.013	.327	.000	.	.875	.737	.008	.555	.003
Partial residual for *stockY*$_2$	Correlation Coefficient	-.189(**)	-.073	-.096	-.144(**)	-.008	1.000	.233(**)	.312(**)	.172(**)	.069
	Sig. (2-tailed)	.000	.155	.063	.005	.875	.	.000	.000	.001	.183
Partial residual for *netbirths*	Correlation Coefficient	-.163(**)	-.143(**)	-.054	-.122(*)	.017	.233(**)	1.000	.659(**)	.582(**)	.356(**)
	Sig. (2-tailed)	.001	.005	.296	.018	.737	.000	.	.000	.000	.000
Partial residual for *netdeaths*	Correlation Coefficient	-.089	-.111(*)	-.092	-.138(**)	-.137(**)	.312(**)	.659(**)	1.000	.395(**)	.497(**)
	Sig. (2-tailed)	.082	.031	.074	.007	.008	.000	.000	.	.000	.000
Partial residual for *BNodes*	Correlation Coefficient	.009	-.094	-.063	-.059	.030	.172(**)	.582(**)	.395(**)	1.000	.508(**)
	Sig. (2-tailed)	.866	.068	.221	.253	.555	.001	.000	.000	.	.000
Partial residual for *DNodes*	Correlation Coefficient	-.144(**)	-.176(**)	-.081	-.041	-.151(**)	.069	.356(**)	.497(**)	.508(**)	1.000
	Sig. (2-tailed)	.005	.001	.116	.422	.003	.183	.000	.000	.000	.
	N	378	378	378	378	378	378	378	378	378	378

Table 6-21 - Spearman correlations between the Schoenfeld partial residuals and the ranked order of survival time (Age) for each covariate - survival analysis of networks.

Notes: ** Correlation is significant at the 1% level (2-tailed);; * Correlation is significant at the 5% level (2-tailed). Only non-censored observations were considered

As we have stated before in Chapter 5, if the correlations obtained with this process are significantly different from zero, and if the corresponding p-value shown in the table as *Sig. (2-tailed)* is lower than 5%, then it will mean that covariates are time-dependent and an extended Cox Model must be computed. The p-value for the correlation test is the p-value for the PH test (in which the null hypothesis is that the PH assumption is not violated). As in Chapter 5, where we have analysed the survival of the firms using a cellular automata approach, the hypotheses for the PH test are as follows:

H_0: the PH assumption *is not* violated,

H_1: the PH assumption *is* violated.

In the first line of Table 6-21 we can see that correlations are significant (p-value<5%) between Age and the covariates identified as *mcost, stockX,* $stockY_2$, *netbirths* and *DNodes*, denoting that the covariate depends on time and therefore the Proportional Hazard Assumption is not supported for that covariate. Since we will continue our Survival Analysis with an Extended Cox Model, we must create time-dependent variables for the covariates in which the PH assumption is not assumed (see Chapter 5). Therefore, we have considered two choices among the most usual forms for defining time-dependent covariates: (Kleinbaum and Klein, 2005)

- the product between the time and the covariate, *TimexCovariate;*
- the product between the natural logarithm of time and the covariate, ln(*Time*)x*Covariate.*

Once again, we used a script in SPSS to create the time dependent covariates[133] for those five covariates in which the correlations with *Age* were significant. The results of the two alternatives for the form of the time dependent covariates are shown in Appendix V and it does not seem to be a great difference between the two choices. Therefore, we have chosen the (*Time)xCovariat*e approach to create time dependent covariates and proceed with the analysis of the *extended* Cox Model. The time dependent covariates are identified as *Covariate(t)*, as, for example, *stockX* (t).

-2 Log Likelihood	Overall (score)		
	Chi-square	df	p-value.
3900,157	223.929	20	.000

Table 6-22 - Omnibus Tests of Model Coefficients in Cox Regression

[133] SPSS does not perform automatically the Cox regression for more than one time dependent covariates simultaneously. Therefore, we created two scripts in *Syntax* mode, a common mode to write programs in a SPSS using a specific language in order to perform more complex computations. Then, we proceeded with the extended Cox model. More details concerning these scripts and the resuts they produced are presented in Appendix V.

The likelihood ratio and chi-square statistics in Table 6-22 are asymptotically equivalent tests of the omnibus null hypothesis that all the coefficients β's are zero. In this case the null hypothesis is strongly rejected.

Covariates	$\hat{\beta}$	SE($\hat{\beta}$)	Wald	df	p-value.	Exp($\hat{\beta}$)
profit	-.035	.017	4.272	1	.039	.966
mcost(t)	2.455	.678	13.119	1	.000	11.651
stockX(t)	.000	.000	10.063	1	.002	1.000
stockY1(t)	-.005	.002	5.551	1	.018	.995
stockY2(t)	.000	.000	2.739	1	.098	1.000
netbirths(t)	-.008	.036	.046	1	.829	.992
netdeaths	.180	.094	3.662	1	.056	1.197
BNodes	.044	.019	5.530	1	.019	1.045
DNodes(t)	-.007	.003	5.631	1	.018	.993
Form			47.832	3	.000	
form(1)	.595	.364	2.673	1	.102	1.814
form(2)	-.472	.372	1.609	1	.205	.624
form(3)	-.721	.389	3.429	1	.064	.487
Strat			10.759	8	.216	
Strat(a)	-.075	.203	.136	1	.712	.928
Strat(b)	.006	.187	.001	1	.976	1.006
Strat(c)	-.295	.211	1.954	1	.162	.744
Strat(d)	.412	.193	4.560	1	.033	1.509
Strat(e)	-.056	.183	.095	1	.758	.945
Strat(f)	-.073	.169	.188	1	.665	.929
Strat(g)	.053	.194	.073	1	.787	1.054
Strat(h)	.197	.189	1.084	1	.298	1.217

Table 6-23 - Summary of covariates statistics in the Cox Regression Equation model[134] for the survival of networks

Note: In the case of categorical covariates *Form* and *Strategy*, the output uses the last category of the variable as reference

[134] As presented above in Chapter 5, column $\hat{\beta}$ contains the estimated coefficients of the covariates; in SE($\hat{\beta}$) the corresponding standard errors are represented; The Wald statistic is computed through the quotient between the coefficient β and its standard error SE(β), and then taking the Normal (0,1) distribution for computing the corresponding value. The column *df* stands for the degrees of freedom involved in the test of hypothesis. The exponentials Exp($\hat{\beta}$) measure the impact of the variables on the hazard of the firms. The column named *p-value* shows the significance of the predictor concerning the test: H_0: $\beta=0$ *versus* H_1: $\beta\neq0$

Table 6-23 shows the coefficients and some corresponding statistics to evaluate the importance of each covariate on the Cox model. According to these coefficients, and to Equation 5-18, the Cox regression model can be written with some of the covariates or predictors as time-dependent variables, as follows:

$$\hat{h}(t,X)=\hat{h}_0(t) \times$$
$$\times e^{-0.035\mathit{profit}+2.455\mathrm{mcost(t)}+0.000\mathit{stockX(t)}-0.005\mathit{stockY1(t)}+0.000\mathit{stockY2(t)}} \times$$
$$\times e^{0.180\mathit{netdeaths}+0.044\mathit{Bnodes}-0.007\mathit{Dnodes(t)}}$$ **Equation 6-41**

The corresponding Cox Regression model is actually a mixed model, as it includes time-dependent and time independent variables[135].

In Table 6-23 *mcost(t), stockX(t), stockY1(t), stockY2(t), netbirths(t)* and *Dnodes(t)* are the six time-dependent covariates that were introduced in the extended model. The overall model is significant ($p \leq 1\%$) and with exception of *Strategy* and *Form*, only *stockY2(t)*, *netdeaths* and *netbirths(t)* have p-values greater than 5%, demonstrating that these covariates have no impact in the survival of firms.

We may use the exponents Exp($\hat{\beta}$) to measure the impact of the variables on the hazard rate of the firms. In fact, Exp($\hat{\beta}$) is the predicted change in the hazard rate for a unit increase in the predictor. Thus, holding the other covariates constant, an additional unit in the contemporaneous density, for example, expressed by *netdeaths* increases the hazard rate by a factor of $e^{+0,180}$=1.197 on average – that is, 19.7 percent. However, according to the Wald statistic, this coefficient ($\hat{\beta}$=+0.180) is not significant (p-value>5%).

The coefficient related to the founding density (*netbirths*) is negative ($\hat{\beta}$=-0.008) and it is also not significant. Therefore, it seems the hypotheses H_{26} and H_{27} are not confirmed.

In what concerns the other covariates, marginal cost, *mcost(t)* has the strongest impact on the hazard rate: it increases the hazard rate by 11 times more. *BNodes* is the other

[135] Time dependent variables are presented with the designation of time *(t)* after their identification.

covariate that has a positive impact in the mortalitity of networks. Other covariates produce the opposite effect on the hazard function: an additional unit of the *profit* decreases the hazard by a factor of $e^{-0,035}$=0.966 (corresponding to almost 4% of decrease, on average). *stockY1, netbirths(t)* and *DNodes(t)* also have negative impact on the mortality of the networks.

Bearing in mind hypotheses H_{26} and H_{27}, we can conclude that they are not confirmed with survival analysis since *netbirths(t)* is found to have a negative impact on the mortality of networks, while *netdeaths* has a positive impact, although there is no statistical evidence to confirm these conclusions.

However, it is worth noting that *BNodes* has a positive impact on the mortality of networks, and simultaneously *DNodes(t)* has a corresponding negative impact. That is, the number of firms at the time of a network's birth has a positive impact on the mortality of networks and the number of firms at the time of a network's death has a negative impact on the mortality of networks. This conclusion is in line with hypotheses H_{210} and H_{211}, which are confirmed.

Considering now the categorical variables, we may conclude that the variable *Form* is important to the survival of networks as we can see from the significance of the test (p-value=0.000). We considered a reference category containing the average values of all categories. Therefore, the other categories are compared to this reference category. We conclude that *Form(1)*, corresponding to linear networks increase the hazard rates of networks, while star and multipolar networks decrease the hazard rate (by almost 40%, and 50%, respectively)[136].

Concerning *Strategy,* it has no overall significant impact on the hazard function. However, considering the comparisons between the reference category and each strategy separately, we may conclude that strategy d) is the only one that shows any significant impact on the hazard rate. In fact, strategy d) increases the hazard rate by more than 50% when compared to the average of all strategies.

---

[136] The impact of the *form* on the survival of networks allows for an affirmative answer to question G6: "Does form matter?". In fact, the shape of the network exerts influence in the mortality of networks.

The following figure shows the cumulative survival functions that are estimated in the analysis. Each collaboration strategy (a to h) is represented by a different colour. These survival curves are useful to describe how the strategies compare over the time period of the study. The survival function is given by:

$$\hat{S}(t,X)=\left[\hat{S}_o(t)\right]^{e^{\sum_{i=1}^{p}\hat{\beta}_i\bar{X}_i}}$$ **Equation 6-42**

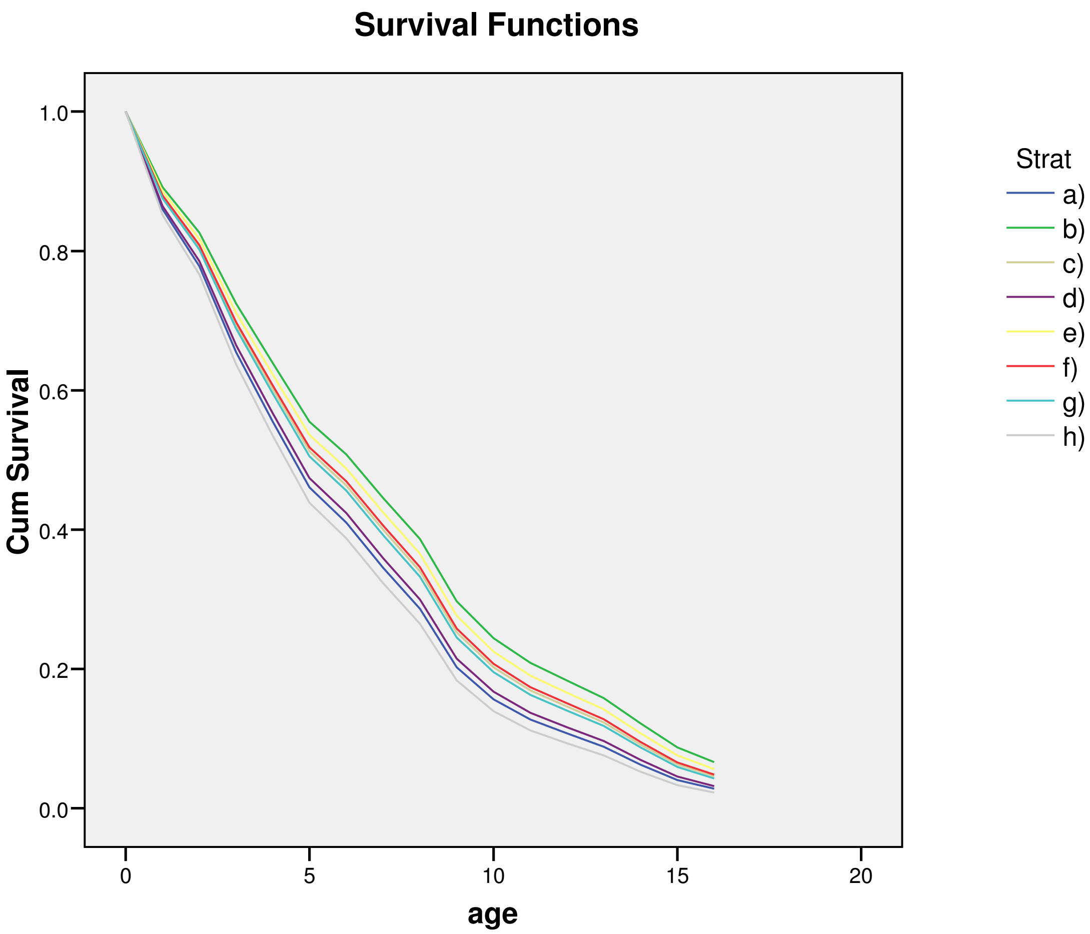


Figure 6-20 – Cumulative survival functions according to the strategies

Note: For visualization purposes, interpolation lines have replaced the original step functions considering all strategies

The variable *age* along the *X* axis in Figure 6-20 represents the age of the networks at the time of their failure and the *Y* axis shows the cumulative survival. The permanent decrease of the lines in the graph is indicative of the fall of the survival of the networks during the twenty generations of their existence. We can see that all strategies have similar cumulative survival paths. However, the values of the cumulative survival functions of strategies b), e), and f) are always higher than others, meaning that networks have higher survival rates, within these strategies.

Survival of firms

Now we will analyse the impact of the covariates on the survival of individual organizations. Two hypotheses are tested within this analysis:

> *H_{25}: Density at founding has a positive impact on the mortality of organizations*
> *H_{26}: Contemporaneous density (of networks) has a negative impact on the mortality of the networks*

A set containing 308 firms was analysed, considering the following covariates:

- Profit (*profit*),
- Marginal Cost (*mcost*),
- *Quantity* (quantity produced by a firm),
- *Links* (the number of links connected to a particular firm),
- Age (*Age*),
- Number of existing nodes (firms) at the time of its birth (*BNodes*),
- Number of existing nodes (firms) at the time of its death (*DNodes*),
- Stock of knowledge in Market X (*StockX*),
- Stock of knowledge in Market Y_1 ($StockY_1$),
- Stock of knowledge in Market Y_2 ($StockY_2$),
- *Status* (dead=0 or alive=1).

BNodes is the covariate used to test H_{25}, while *DNodes* is the covariate use to test H_{26}. We repeated the previous Proportional Hazard (PH) test and started by computing the Spearman correlations (since the partial residuals do not follow the Normal distribution) between the Rank of Age and the partial residuals corresponding to the covariates.

		Rank of *Age*	Partial residual for log*Profit*	Partial residual for *mcost*	Partial residual for *Quantity*	Partial residual for *Links*	Partial residual for *BNodes*	Partial residual for *DNodes*	Partial residual for *StockX*	Partial residual for *StockY1*	Partial residual for *StockY2*
Rank of *Age*	Correlation Coefficient	1.000	-.578(**)	-.088	-.943(**)	.097	.156(**)	.315(**)	-.759(**)	-.277(**)	-.249(**)
	Sig. (2-tailed)	.	.000	.121	.000	.091	.006	.000	.000	.000	.000
Partial residual for log*Profit*	Correlation Coefficient	-.578(**)	1.000	-.020	.654(**)	.143(*)	-.227(**)	-.282(**)	.413(**)	.186(**)	.104
	Sig. (2-tailed)	.000	.	.721	.000	.012	.000	.000	.000	.001	.070
Partial residual for *Mcost*	Correlation Coefficient	-.088	-.020	1.000	.001	-.258(**)	.030	-.127(*)	-.030	-.098	-.244(**)
	Sig. (2-tailed)	.121	.721	.	.992	.000	.597	.026	.597	.085	.000
Partial residual for *Quantity*	Correlation Coefficient	-.943(**)	.654(**)	.001	1.000	-.086	-.157(**)	-.315(**)	.718(**)	.245(**)	.215(**)
	Sig. (2-tailed)	.000	.000	.992	.	.134	.006	.000	.000	.000	.000
Partial residual for *Links*	Correlation Coefficient	.097	.143(*)	-.258(**)	-.086	1.000	.027	.303(**)	.103	.212(**)	.164(**)
	Sig. (2-tailed)	.091	.012	.000	.134	.	.640	.000	.071	.000	.004
Partial residual for *BNodes*	Correlation Coefficient	.156(**)	-.227(**)	.030	-.157(**)	.027	1.000	.738(**)	-.057	.036	.074
	Sig. (2-tailed)	.006	.000	.597	.006	.640	.	.000	.321	.530	.194
Partial residual for *DNodes*	Correlation Coefficient	.315(**)	-.282(**)	-.127(*)	-.315(**)	.303(**)	.738(**)	1.000	-.077	.064	.126(*)
	Sig. (2-tailed)	.000	.000	.026	.000	.000	.000	.	.177	.260	.027
Partial residual for *StockX*	Correlation Coefficient	-.759(**)	.413(**)	-.030	.718(**)	.103	-.057	-.077	1.000	.210(**)	.308(**)
	Sig. (2-tailed)	.000	.000	.597	.000	.071	.321	.177	.	.000	.000
Partial residual for *StockY1*	Correlation Coefficient	-.277(**)	.186(**)	-.098	.245(**)	.212(**)	.036	.064	.210(**)	1.000	-.326(**)
	Sig. (2-tailed)	.000	.001	.085	.000	.000	.530	.260	.000	.	.000
Partial residual for *StockY2*	Correlation Coefficient	-.249(**)	.104	-.244(**)	.215(**)	.164(**)	.074	.126(*)	.308(**)	-.326(**)	1.000
	Sig. (2-tailed)	.000	.070	.000	.000	.004	.194	.027	.000	.000	.
	N	308	308	308	308	308	308	308	308	308	308

Table 6-24 - Spearman correlations between the Schoenfeld partial residuals and the ranked order of survival time (Age) for each covariate - survival analysis of firms

We observe in Table 6-24 that the covariates having significant correlations with time are: *profit*, *Quantity, BNodes, DNodes, stockX, stockY1* and *stockY2*. Therefore, we created time dependent covariates *profit(t)*, *Quantity(t), BNodes(t), DNodes(t), stockX(t), stockY1(t)* and *stockY2(t)* [137], and obtained the following results in the extended Cox model:

-2 Log Likelihood	Overall (score)		
	Chi-square	df	p-value.
3098.633	289.153	9	.000

Table 6-25 - Omnibus Tests of Model Coefficients in Cox Regression

[137] As before, we have chosen the (Time)xCovariate approach to create time dependent covariates, using a specific script in SPSS.

The model is significant (p-value<1%) and the impact of each of the covariates on the mortality of firms is identified in Table 6-26.

Covariates	$\hat{\beta}$	SE($\hat{\beta}$)	Wald	df	p-value	Exp($\hat{\beta}$)
profit(t)	-.014	.011	1.693	1	.193	.986
mcost	.569	5.652	.010	1	.920	1.766
Quantity(t)	.000	.000	.199	1	.656	1.000
Links	-.281	.031	79.835	1	.000	.755
BNodes(t)	-.055	.009	34.108	1	.000	.947
DNodes(t)	.050	.005	91.326	1	.000	1.052
stockX(t)	.000	.000	1.984	1	.159	1.000
stockY1(t)	.001	.001	3.094	1	.079	1.001
stockY2(t)	.001	.000	2.115	1	.146	1.001

Table 6-26 - Summary of covariates statistics in the Cox Regression Equation model for the survival of firms

According to the p-values associated with the Wald statistic, we can conclude that only three covariates have a significant impact (p-values<1%) on the firms' hazard rate: *Links, BNodes(t) and DNodes(t)*. The Cox regression model can be written as follows:

$$\hat{h}(t,X)=\hat{h}_0(t) \times e^{-0.281\,Links-0,055\,Bnodes(t)+0,05\,Dnodes(t)}$$ **Equation 6-43**

The number of links (*Links*) clearly decreases the hazard rate of the firms: the increase of one link decreases the hazard by 24.5%. The number of firms at the time of birth (*BNodes*) has a negative impact on the hazard rate, and the number of firms at the time death (DNodes) has a corresponding positive impact. This fact contradicts the initial hypotheses H_{24} and H_{25} that are not confirmed by suirvival analysis.

In the next section, we search for temporal patterns that are related to the evolution of networks. Techniques of Multivariate Data Analysis are used, such as Multiple Factorial Analysis and the STATIS Method.

6.9.5. Clustering the evolution of networks

The aim of this section is to detect time clustering of networks and variables. The techniques that are used to perform such analysis fit in the methods of Temporal Data Mining or Evolutionary Data Analysis. In this study, such methods provide an underlying structure of networks having similar performance, and identify the variables that contribute most to the structure design. We start by presenting the database and proceed with STATIS Dual method that aims at describing the inter-structure and intra-structure of the data. Finally a Multiple Factorial Analysis followed by a clustering meythod is applied in order to identify the main variables that characterize the patterns of network evolution.

Data

Data was generated by the simulation described in previous sections. Each observation (network) is characterized by a set of four attributes: profit (V_1), stock of knowledge of market X (V_2) [138], marginal cost (V_3) and network diameter (V_4). In this kind of analysis we build a data table D for every time step, where n observations at each time period, t, are described by a group of variables, V={V_1, …V_4}, as presented in Figure 6-21.

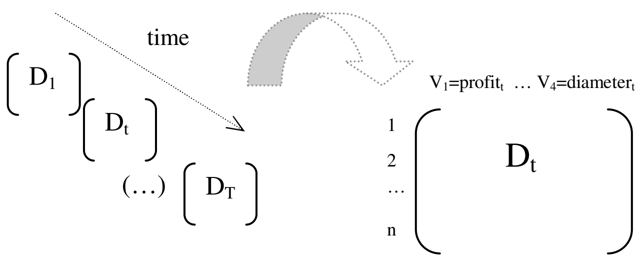


Figure 6-21 – Representation of the data sets to be analysed

Twenty original time periods of the simulation, corresponding to the generations of the evolutionary process, were aggregated in five time periods, in order to simplify the

[138] For this analysis only the stock of knowledge concerning market X considered, since it represents the most representative values.

analysis[139] (each aggregation containing the arithmetic mean of the corresponding variables within each strategy). Therefore, concerning the variable profit (V_1) and the network i, at time step t_1, (aggregating periods 1 to 4), represented by the element D_{i,t_1,V_1}, the temporal aggregation, is defined by:

$$D_{i,t_1,V_1} = \frac{\sum_{t=1}^{4} \text{profit}_{i,t}}{4} \qquad \textbf{Equation 6-44}$$

Consequently, five different periods (T_1, …, T_5) characterized by four variables in each, in a total of 5x4=20 variables were created. We have gathered 61 observations, each of which corresponds to an organizational network[140].

	V1	V2	V3	V4
Period 1 (T_1) (original time periods 1 to 4)	Profit1	Stock1	MargC1	Diameter1
Period 2 (T_2) (original time periods 5 to 8)	Profit2	Stock2	MargC2	Diameter2
Period 3 (T_3) (original time periods 9 to 12)	Profit3	Stock3	MargC3	Diameter3
Period 4 (T_4) (original time periods 13 to 16)	Profit4	Stock4	MargC4	Diameter4
Period 5 (T_5) (original time periods 17 to 20)	Profit5	Stock5	MargC5	Diameter5

Table 6-27 – Structure of the matrix where the variables are captured in five different time steps

[139] Since we are searching for evolutionary paths, it is preferable to use five time periods in the analysis than twenty (the complexity arises with the number of time periods). Therefore, each new period corresponds to the aggregation of four original time periods.

[140] In each period of time we considered the same set of observations. Therefore, the n=61 observations (the networks) were organized according to the collaboration strategies in the following way:

Network	Strategy
1 to 9	1
10 to 16	2
17 to 24	3
25 to 29	4
30 to 40	5
41 to 46	6
47 to 51	7
52 to 60	8

In order to study the evolution of the networks and to discover possible similar growth paths, we started with a Time-Dependent Multivariate Analysis (using the *STATIS Dual* method), complemented by a Principal Component Analysis to follow the trajectories of the variables. A Multiple Factorial Analysis has also been performed in order to obtain a most clear clustering of the observations.

STATIS Dual

The STATIS [141] method (Escoufier and Pagés, 1988; Lavit, 1988; L'Hermier des Plantes, 1976; Dazy and Le Barzic, 1996) is suitable for exploring several data tables indexed by time simultaneously. It applies to quantitative data collected in one of the following situations:

- *T data tables collected in different occasions for the same set of observations; variables can be different;*
- *T data tables collected in different occasions for the same set of variables; observations can be different.*

These two situations correspond to different strategies: the first one underlines the proximities among observations (STATIS), while the second one emphasizes the relationships among variables (STATIS dual), and that is the one we will apply in our analysis.

Information was gathered for the same variables in different periods of time, and the number of individuals is the same in every period. Information was organized in such a way that the T data tables stay side by side in the data matrix. This information has been registered in a data matrix having the following format (considering four variables gathered in five time groups with the same n individuals in each):

[141] The acronym STATIS stands for: *Structuration de Tableaux à Trois Indices de la Statistique* (Structuring of Three Index Tables of Statistics).

	T_1			T_2			(…)			T_5		
	V1	(...)	V4	V1	(...)	V4	V1	(...)	V4	V1	(...)	V4
1												
2												
(...)												
n												

Table 6-28- Structure of the Matrix for the *Statis Dual* approach

In what follows, we will analyse the *inter-structure* and the *intra-structure* of the data. These two steps correspond to two fundamental steps in the STATIS (and STATIS Dual methods). In the former, our goal is to measure the distances between the tables, i.e., to study the evolution of the time periods that characterize the organizational networks; in the latter we will analyse the association between the variables that belong to the same period of analysis, and find an optimal plan for the representation of the variables measured in different periods[142].

Data Inter-structure

The analysis of data inter-structure corresponds to the study of the relationships between different time periods, given by tables T1 to T5. The Euclidean distances between the five correlation matrices, one for each period were computed and used to achieve this goal (see Table 6-29). With this information, we can evaluate the evolution of the networks along five time steps.

	T1	T2	T3	T4	T5
T1	0.000				
T2	0.464	0.000			
T3	0.638	0.727	0.000		
T4	0.839	0.993	0.380	0.000	
T5	2.210	2.525	1.970	1.682	0.000

Table 6-29 – Euclidean distances between the five different correlation matrixes

[142] To perform the multivariate data analysis in this section (STATIS and Multivariate Factor Analysis), we used the statistical software *SPAD (Système Portable d'Analyse de Données, Decisia,* 1996-2007).

It is clear that distance between matrices increases with time. For example, the distance between T1 and T2 (0.464) is lower than the distance between T2 and T3 (0.767) and so on. A possible explanation for this behaviour is that networks change with time and that this change is greater along the periods.

Table 6-30 shows the eigenvalues of the inter-structure matrix [143] after a Principal Component Analysis. The weight of each eigenvalue defines the relative importance of the corresponding axis. The column *percentage* in this table represents the weight of the corresponding eigenvalue in the sum of the eigenvalues and it is a measure of the importance of the axis (also known as *component* or *factor*) associated with the eigenvalue. The column *cumulated percentage* represents the total of explained variance of the data.

Eigenvalue	Value	Percentage	Accumulated percentage	Histogram
1	40.554	92.47	92.47	**********************************
2	0.268	6.11	98.58	***
3	0.0387	0.88	99.46	*
4	0.0236	0.54	100	*
5	0	0	100	*

Table 6-30 – Eigenvalues of the *Statis Duale* inter-structure analysis

We can observe that the first axis explains more than 90% of the total variance of the data. Representing now the five data tables in the system of coordinates built on the plane[144], we will get the following graph (the edges have been added in order to give a sequence to the evolution of the tables – from table T1 to table T5):

[143] The inter-structure matrix contains the Hilbert-Schmidt scalar-products between tables Ti and Tj (i,j=1,2,…,5, with i≠j), (Dazy and Le Barzic, 1996).
[144] The plane refers to the system of axis corresponding to the first two eigenvalues. This plane is also known as the first factorial plane because it is formed by the axes 1 and 2. In this type of representation, the coordinates of each point are the correlations between the variable and the corresponding axis. The terms axis, factor and component are used as synonyms.

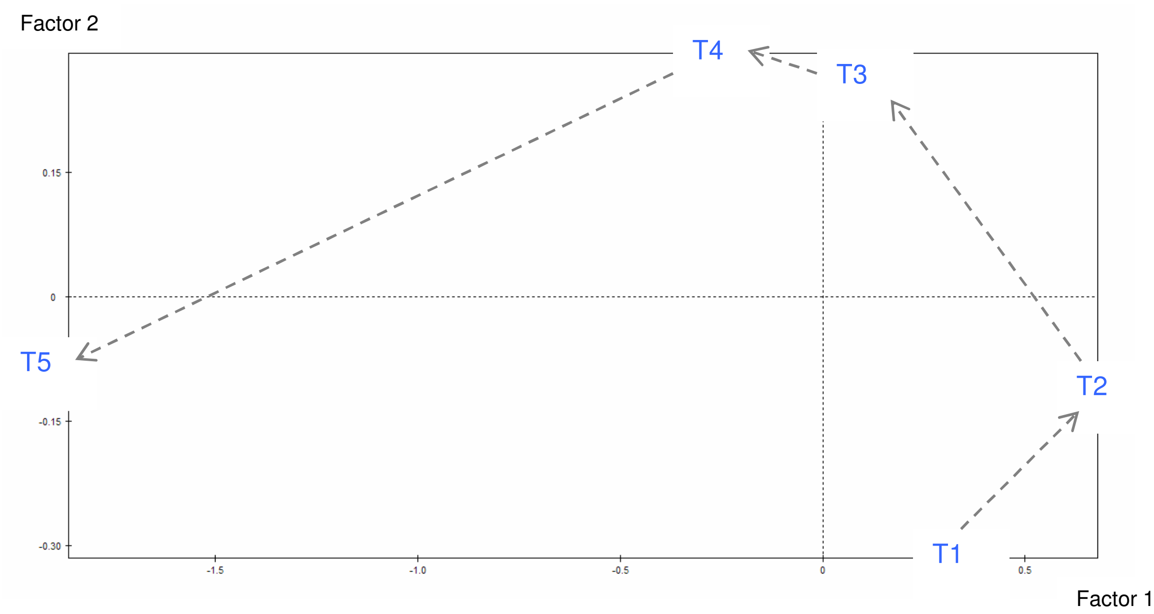


Figure 6-22- Data inter-structure corresponding to the evolutionary path of Ti matrices.

In Figure 6-22, Factor 1 and Factor 2 represent the axes associated respectively with the first and second eigenvalues. Factor 1 discriminates tables T1, T2 and T5 from tables T3 and T4[145]. After computing the distances between the several tables and the average table (*compromise* table) we conclude that tables T3 and T4 are close to the *compromise*, while T1, T2 and T5 are distant from the *compromise*. On the other hand, Factor 2 clearly distinguishes T5 from the rest of the other factors. The period T5 corresponds to last generations of the simulation (17 to 20), in which network attributes get the highest values. That is why T5 is represented distant from the other tables.

Data Intra-structure

[145] One step in STATIS methods that was not defined before is the definition of a *compromise* that aims to summarize the five T tables into one only, named *compromise table*. This table aims at being representative of the whole set of tables. This compromise is a weighted mean of the observations in each matrix Ti (i=1, 2, …, 5) and is defined by a matrix of *compromise* correlations. In our case, this is a 4x4 matrix containing the correlation between the four variables.

	V1	V2	V3	V4
V1	1			
V2	0.586	1		
V3	0.360	0.465	1	
V4	0.703	0.467	0.441	1

The analysis of data intra-structure takes into account the correlations of the variables V_1 to V_4 and the representation of the observations (networks 1 to n), characterized by the whole set of T tables. The *compromise* table defines the correlations between variables (see footnote 145). A representation of the variables is shown in Figure 6-23 in which the correlations between each principal component of the *compromise* table and each variable are represented through the STATIS method. Proximities between variables within this representation denote strong correlation between them. In this figure, the correlations between variables V1 (profit), V4 (diameter) and V2 (stock of knowledge) are higher than with V3 (marginal cost).

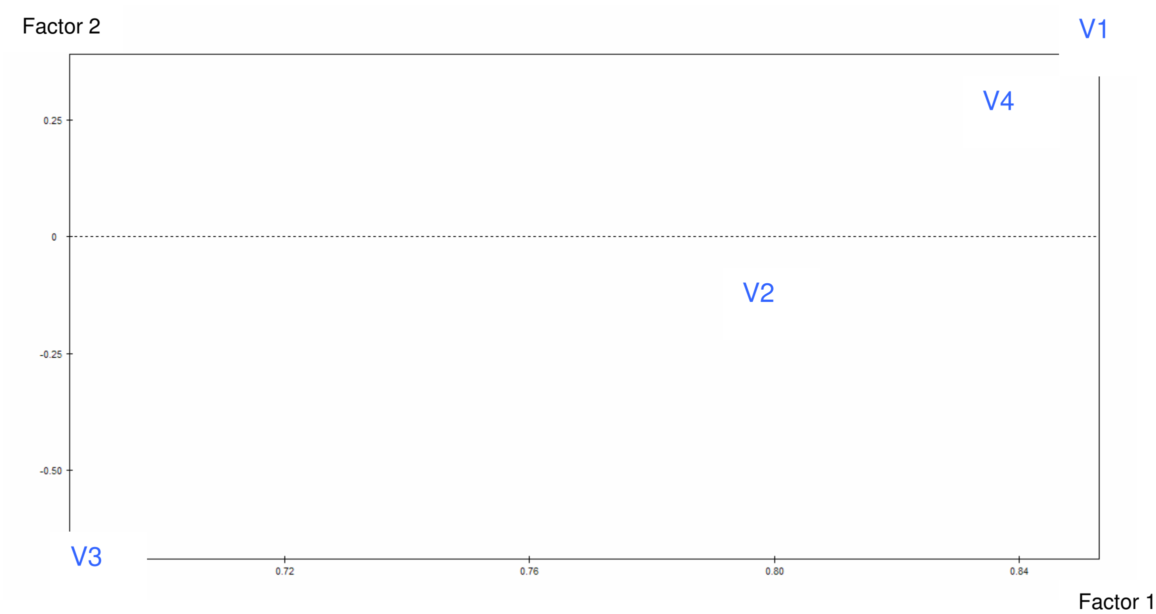


Figure 6-23 – Representation of variables in the plan defined by Factor 1 and Factor 2 of the *compromise table* capturing data intra-structure

In order to represent the evolution of the variables throughout the five periods of time, we have performed a Principal Component Analysis of the table composed by the five juxtaposed tables corresponding to the periods of time. The SPAD package offers the possibility of representing the trajectories of the variables corresponding to the five periods of analysis, by depicting the variables at time 1 and using the variables related to times 2 to 5 as supplementary variables. The factorial plane of this analysis is optimal for representing the set of points of the trajectory. Once again, the coordinates of each variable in this plane are defined by the correlations between the variable and the corresponding axis.

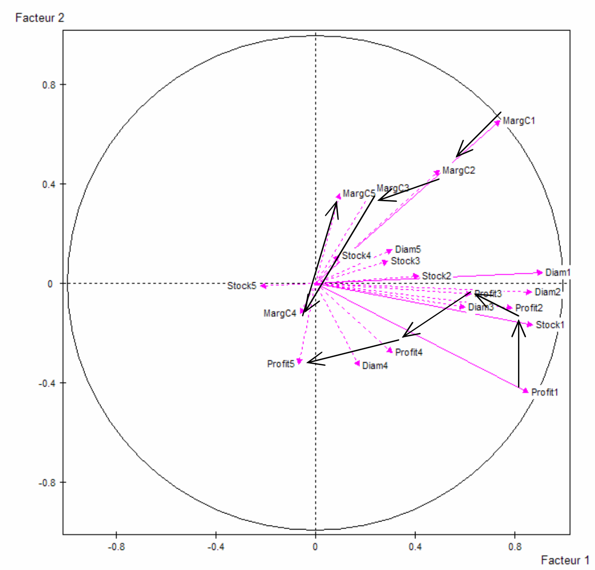


Figure 6-24 – Trajectories of the variables represented in a PCA factorial plane

Note: only the trajectories of V1 (profit) and V3 (Marginal cost) are represented in the graph.

Taking into consideration the correlations between each variable and the corresponding axis or factor (see Appendix VI) we can state that the factor 2 is related to the marginal cost (V3) since the correlations are greater with this variable. On the other hand, factor 1 represents the rest of the variables (profit, stock of knowledge and diameter), in the first, and in the succeeding periods. Therefore, the evolution of the trajectories is characterized by two different behaviours, according to the variables being analysed: marginal cost (that decreases, in average, along time) and the other variables (that increase, in average, along time).

We can do a similar analsys considering the observations (the networks) instead of the variables. However, we decided to perform a complementary type of analysis (*Multiple Factorial Analysis*) followed by a clustering of the networks in order to better discriminate the networks.

Multiple Factorial Analysis (MFA)

MFA (Escoufier and Pagés, 1998; Dazy and Le Barzic, 1996) is appropriate when we have the same number of variables and observations measured in different times. Although this is not the case (our observations, the networks are not the same in all the periods), such analysis is still relevant, because of the possibility of clustering the final results. Considering that observations (rows) will be filled with zeros in the data matrix for the times (groups of variables) where they were not observed, the AFM method can still be a good way to analyse the evolution of the networks.

The data base is the same as in STATIS situation. We used the software SPAD again to perform the MFA, followed by a hierarchical clustering on the obtained factors. After the clustering step we have obtained the dendrogram shown in . Two different relevant partitions (in 2 and in 4 classes) are depicted.

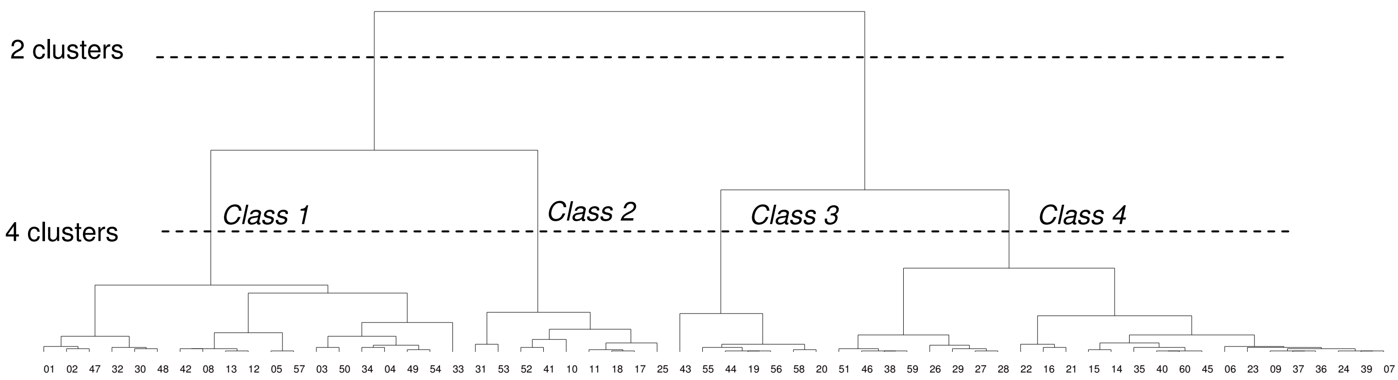


Figure 6-25 - Dendrogram of the clustering

The best partition (that discriminates better the networks in the study) is the one with four classes (according to the agglomeration schedule, provided by the output of SPAD - not given here). Table 6-31 distinguishes two groups of networks, according to the classes to which they belong (considering the partition in four classes).

Classes	Class 1	Class 2	Class 3	Class 4
Networks' ID	1, 2, 47, 32, 30, 48, 42, 8, 13, 12, 5, 57, 3, 50, 34, 4, 49, 54, 33,	31, 53, 52, 41, 10, 11, 18, 17, 25	43, 55, 44, 19, 56, 58, 20	51, 46, 38, 59, 26, 29, 27, 28, 22, 16, 21, 15, 14, 35, 40, 60, 45, 6, 23, 9, 37, 36, 24, 39, 7
Number of networks per class	19	9	7	25

Table 6-31 – Identification of the networks according to the classes after the clustering method

A question arises of what this aggregation in four classes discriminates in terms of the main variables that characterize the networks. SPAD provides a list of the variables that contribute most for each class, according to the partition in four classes. Table 6-32 presents a summary of these contributions[146]. Classes 2 and 4 are associated with similar sets of variables, but the means of the variables in Class 4 are lower than the general means, while the means of the variable in class 2 are higher (see Appendix VI for details). Networks of Class 3 are associated with great values of stock of knowledge (in periods T3 and T4) and great diameter. Concerning Class 1, networks are characterized by high values of Marginal cost, and low values of profit and diameter.

[146] After the clustering of the networks, it is possible to identify what are the variables that contribute most for each class. These variables (named characteristic variables) are those in which the difference between the mean of each class and the general mean is statistically significant, according to a particular statistical test (Escoufier et Pagés, 1998). A more detailed description of these classes and applied tests is provided in Appendix VI.

		Class 1	Class 2	Class 3	Class 4
T1	V1 (Profit)		X		X
	V2 (Stock Knowl.)		X		X
	V3 (Marg. Cost)		X		X
	V4 (Diameter)		X		X
T2	V1 (Profit)		X		X
	V2 (Stock Knowl.)	X	X		X
	V3 (Marg. Cost)	X			X
	V4 (Diameter)		X		X
T3	V1 (Profit)		X	X	X
	V2 (Stock Knowl.)		X		X
	V3 (Marg. Cost)	X			X
	V4 (Diameter)		X	X	X
T4	V1 (Profit)	X	X		
	V2 (Stock Knowl.)			X	X
	V3 (Marg. Cost)				
	V4 (Diameter)	X	X		
T5	V1 (Profit)				
	V2 (Stock Knowl.)			X	
	V3 (Marg. Cost)	X			X
	V4 (Diameter)				X

Table 6-32 – Importance of the variables to the class formation (characteristic variables are marked with a X)

Looking at the networks (observations) that belong to each one of those classes, we may establish a correspondence between classes and strategies, by building a contingency table as follows:

Strategy	Class 1	Class 2	Class 3	Class 4
a)	6			3
b)	2	2		3
c)		2	2	4
d)		1		4
e)	4	1		6
f)	1	1	2	2
g)	4			1
h)	2	2	2	2
total	***19***	***9***	***7***	***25***

Table 6-33 – Number of networks according to the classes of the clustering and collaboration strategies

It can be seen that networks fall in different classes and that there is not a clear pattern that can be related to the collaboration strategies. Class 4 (related to lower values in all the variables) contains networks of all strategies and is difficult to identify any association with strategies.

However, the other classes make clear some kind of association with collaboration networks: Class 1 is related to higher marginal costs that can be confirmed by its major association with strategy a), g) and e) (see Table 6-9). Class 2 is represented by several networks from strategies b), c) and h), with high average profit and diameter. Class 3 contains networks from strategies c), f) and h) that are among the strategies with higher average levels of stock of knowledge.

Therefore, looking at the networks (observations) that belong to each one of those classes, it seems that the clustering method has opposed the networks with higher levels of stock of knowledge from those with higher values of profit and diameter and from those with high marginal costs.

6.10. Concluding remarks: interpretation and systematization of the results

The results presented in the previous sections allow us to draw several conclusions about our model. Observations of others, based on empirical studies, were taken into account in the development of the model, and therefore should be reproduced in the results. In the next paragraphs we summarize the main results and the main conclusions that we can obtain from NetOrg.

We have analysed and compared the outputs of the model after running different collaboration strategies. Strategies b) (*Average Stock Scenario with Great Expected Value)*, and d) (*Concentration process)*, are the most profitable, on average[147]. Profit is

[147] In our work, central nodes of the networks following strategy d) (concentration process) show great values of profit and stock of knowledge. This effect is comparable to the "*rich gets richer*" paradigm of

also high in Strategy c) (*Preferential meeting process with Profit evaluation*). Simultaneously, marginal costs are lower, on average, in these three strategies b), c), and d), being higher in strategies a) and g).

Concerning the network properties, we found an association between transitivity (the cluster formation in networks) and profit: the higher the transitivity, the higher the profit of the networks, in average. The *concentration process* is considered the strategy with higher transitivity and density.

We have found that profit is lower in conservative firms (the resulting p-value was lower than 1%). The network support, another cognitive attribute, produces an significant impact on marginal cost and on the stock of knowledge: marginal costs are lower and stock of knowledge are higher when *network support* is activated. On the other hand, no significant impact is observed on profit.

The effects of learning are not significant for the increasing of the profit, concerning the population of networks. However, results differ, if we analyse the population of firms instead of networks. In this case, learning significantly increases the stock of knowledge of networks and slightly increases their profit although this increase is not significant.

The effects of migration to regions with lower marginal costs make the firms increase their profits, because of the reduction of their marginal costs. Only firms with positive attitude to risk migrate and therefore these firms tend to perform better than the conservative ones.

As expected, the spread of innovation is one of the main effects of the networks. Firms in networks have higher life expectancies (more than half - when compared to the non-networked firms) and their profit is considerably superior.

In what concerns the technological distance between two firms (not the geographical one), we have concluded that, on one hand if it is short they tend to cooperate horizontally, and if it is high, they tend to cooperate vertically. On the other hand, for

---

Barabasi (2002). According to this author, studies of the evolution of system structures showed that there is a "preferential attachment" to existing nodes leading to a "winner-take-all" outcome.

Technological distances the opposite is true: if the distance between two firms is high, they tend to cooperate horizontally; if the distance between two firms is short, they tend to cooperate vertically. We have also concluded that firms having long-term relationships liver longer in average.

Regarding the survival analysis, and the significance of the Cox Regression coefficients, we have concluded that contemporaneous density has a positive impact on the mortality of the networks while density at founding has a negative impact. Therefore, the hypotheses H_{26} and H_{27} a not confirmed. Profit decreases the hazard rate and marginal cost increases it by eleven times more. Simultaneously, the number of firms at the time of a networks birth has a positive impact on the mortality of networks and the number of firms at the time of a networks death has a negative impact on the mortality of networks. Therefore hypothesis H_{210} and H_{211} are confirmed.

The form of the network is relevant to the survival of networks as can be observed from the significance tests: The linear form increases the hazard rate, while star and multipolar forms strongly decreases the hazard rate by 40% and 50% respectively.

The type of strategy has no overall important impact on the hazard function with the exception of Strategy d) *(concentration process),* which holds one of the highest average profits but it increases the hazard rate by more than 150%.

The following table summarizes the conclusions reached in this chapter:

Specific Hypothesis	Empirical evidence (summary)	General Question	Result
H_{21}	Buyers promote both a concentration process of component suppliers and room for creating important collaborations and alliances with suppliers and among suppliers and buyers; Networks are formed around buyers and several topologies are observed.		**Confirmed**
H_{22}	Some cooperative firms have increased their profits because of the reduction of costs due to the rationalization of some of their functions [through cooperation via network formation].	G4	**Confirmed**
H_{23}	Migration of firms to markets with lower marginal costs is as a way to rationalize production		**Confirmed**
H_{24}	Contemporaneous density has a negative impact on the mortality of organizations	G1	Not Confirmed
H_{25}	Density at founding has a positive impact on the mortality of organizations	G1	Not Confirmed
H_{26}	Contemporaneous density has a negative impact on the mortality of the networks	G2	Not Confirmed
H_{27}	Density at founding has a positive impact on the mortality of the networks	G2	Not confirmed
H_{28}	Supplier firms that have long-term relationships with their buyers will live longer, in average	G7	**Confirmed**
H_{29}	If the distance between two firms is short, they will have a higher tendency to cooperate horizontally; if the distance between two firms is high, they will have a higher tendency to cooperate vertically	G8	**Confirmed (for thechnological distance)**
H_{210}	Contemporaneous density (of firms) has a negative impact on the mortality of networks	G3	**Confirmed**
H_{211}	Density at founding (of firms) has a positive impact on the mortality of network	G3	**Confirmed**

Table 6-34 - Summary of the tests that were made in Chapter 6

Finally, with Multiple Factorial Analysis and the *Statis* methodology, we have analysed the temporal evolution of the networks and concluded that the clustering method has opposed the networks with higher marginal cost to those with high levels of the stock of capital and to those with higher values concerning other variables, such as profit and diameter. Networks were clustered considering these differences, and methods of evolutionary data analysis suggest that the stock of knowledge is one of the main features of the networks in NetOrg..

In this work we observe that organizations increase their stock of knowledge due to the spread of innovation and that is a direct consequence of the fact that organizations are linked to networks. The network is a mechanism undergoing adaptation[148], where the internal synergies shape the course of evolution.

We recall, as a conclusive remark, the phrase of Edgar Morin that configures our definition of network as an organization: *an organization ensures a high degree of interdependence and reliability, thus providing the system with the possibility of lasting for a certain length of time, in spite of chance disruptions*.

[148] No individual fitness measures were defined in NetOrg, although the profit of a firm or network can be seen as an indicator of the individual or collective adaptation.

7. Conclusions

Final remarks

In this work, we have analysed the interactions among organizations and between them and the environment, with the aim of studying:

- the survival of organizations based on density dependence;
- the legitimation limits of a population of organizations;
- the way how cooperation motivates the formation of networks
- the spread of innovation through organizational networks;
- the effects of networking on the survival of organizations;
- the effect of networking strategies on the overall profit and other variables;
- the impact of cognitive capabilities on the survival and profit;

Our main perspective of analysis is based on the works of organizational ecologists as well as on firm' Demographers', in which organizational populations are seen as a whole. Ecological concern is most heavily concentrated on the probability of survival of the organizational form, not the individual organization. In this sense, we are much more interested in studying the evolutionary properties and the evolution of the populations rather than the individual firms behaviour.

However, individual adaptation is determinant to the improvement of a group. In contrast with the majority of the works related to the Ecology of Organizations, we were concerned about individual decisions and the resulting impact on the whole population.

This perspective is in line with mainstream economics, who use to focus on the individual firms. They attribute rational behaviour to firms in addition to profit motives. Individual decisions taken in organizations are commonly based in individual fitness measures, related to economical goals that are not captured by ecological models.

For that reason we have defined a individual, *microeconomic* modelling and a set of individual attributes in order to model the behaviour of an individual organization and to study the impact of that behaviour on the whole population.

Within this individual *bottom-up* perspective, where the firm is the centre of the model configuration, we have introduced *collaboration* and *cooperation*. These nuclear concepts appeared as a need to complement the action and the capacities between organizations, namely in terms of technological innovation. Inter-organizational networks are a direct result of this aggregation between firms.

Space and location were considered important components for the study of the organizational survival as the organizational environments have spatial components that affect the evolutionary dynamics of organizational populations. Each population of organizations occupies a different niche and the size of this niche may be determinant in a density dependence perspective.

Therefore, we have developed two case studies using a simulation-based methodology. In the first study, a Cellular Automata-based application, CASOS, was implemented. Our main objective with this application was to analyse the effects of the contemporaneous and founding density on the mortality of organizations and, simultaneously, to determine the density interval limits of a population, a niche, that influence the survival and the founding of firms.

We have used a Genetic algorithm to calibrate the simulation and therefore approximate the simulated results to real data. We have concluded that the system performance is rather consistent with the reality observed from collected data in what concerns the total number of firms and the birth rate, but the simulated death rates that intervene in the Fitness function of the calibration process differ much from the real ones.

However, the final solutions for the values of the density interval limits, Size and Age appear to be stable if intervals are considered. They have been found to be the following:

DSl={3:5}; DSu={10:28}; DBl={2:4}; DBu={14:25}; Sl={1:4};
Al={1} Equation 5-21

This roughly means that birth density interval limits are wider than survival density intervals. New firms are almost always accepted but only some of them will survive:

infant firms are most likely to fail particularly when they are born (the Age limit is 1) and firms that attain the age of 1 tend to prevail. The size limit is more tolerant: the higher limit S_l=4 makes part of the solution as the size limit for a firm to be considered legitimated.

The density survival (DS) levels, measured in number of industries per square kilometre (x10), lay roughly between 3 and 28, for the period of analysis, constituting a type of legitimation limit implicitly established in the region. Within this point of view, this organizational population would reach the top of its legitimation at the upper bound Density Survival, i.e., a value close to 28. Actually, the real value of DS_u for the region of Ave is 25. According to organizational ecology theories, we believe that the density survival upper limit has been reached in the region of Ave, and therefore the density will decrease fast.

These conclusions can help in the definition of political measures to improve the birth of new firms in the region, and to avoid the disappearance of others.

The signs of the covariates obtained with the survival analysis also confirm the capability of the model to reproduce the reality: the effect of the *size* on firm's mortality is negative and statistically significant, confirming what was said in literature. The sign of the covariate *contemporaneous density* is also negative as in literature, but with no significance.

In a second study, NetOrg, a Multi-Agent framework was developed to analyse the dynamics of organizational survival in cooperation networks. According to evolutionary theories, firms innovate in order to increase their survival rates and the process of innovation is related to the creation of networks of firms. So, the main goal here was to study the impact of individual attributes on the networks that emerged through the implementation of different collaboration strategies.

In this second case, interactions between organizations are more complex, given that the framework includes a set of initial definitions, a microeconomic model, a decision-making process, and a cognitive model.

Set of initial definitions:

- *technological, physical distances and a total distance;*
- *model of knowledge creation*

The *microeconomic* modelling includes:

- *market demand;*
- *production costs;*
- *profit;*

The process of decision making considers:

- *creation and spread of innovation;*
- *collaboration/Cooperation;*
- *contracts:*
- *organizational survival:*
- *product quantities;*
- *migration.*

The cognitive model includes:

- *Learning;*
- *Risk;*
- *Network support.*

To validate our modelling approach, we have considered some empirical evidence and therefore have decided to focus our analysis on three real life examples from the areas of Automobile manufacturing, Textile Industry and e-Marketplaces. The main empirical evidences that we took from real-life examples are part of our research hypotheses and have been used to guide the validation of our model. Other hypotheses that were tested served as a way to extract knowledge produced by the model.

The main conclusions that we have drawn from NetOrg are the following:

- Only firms with high values of risk are willing to migrate; these firms seem to perform better than conservative ones. The effects of migration to regions with lower marginal costs allow firms to increase their profits, because of the reduction of their marginal costs.

- The effects of learning are not significant if we consider the population of networks. However, in a firm perspective, it tends to increase significantly the stock of knowledge of networks (question G5)
- Firms in networks have higher life expectancies compared to firms that re not in networks and their profit is considerably higher (question G4).
- It can also be said that firms having long-term relationships liver longer in average (question G7)
- Regarding the survival analysis, and the significance of the Cox Regression coefficients, we have concluded that Contemporaneous Density has a positive impact on the mortality of the networks while density at founding has a negative impact. Marginal cost increases drastically the hazard and the profit reduces it the by 4% (question G3)
- We have also concluded that if the distance between two firms is short, they tend to cooperate horizontally and if the distance between two firms is high, they tend to cooperate vertically (question G8).
- The Form of the network is relevant to the survival of networks as stated by the significance of the test. The hazard rate is increased with the linear form, and it decreased with star and multipolar forms. We concluded that the lattest forms support the survival of firms (question G6).
- Strategy choice has an important impact on the performace of firms: strategies b), and d), are the most profitable of all. Marginal costs are higher in strategies a) and g)
- The diameter of networks decreases along time and transitivity (the clustering coefficient of networks) tends to increase strongly in the initial life of neworks and decreases slowly at the end.

Concerning questions G9 ("*How does knowledge disseminate through networks?*"), and G10 (*"How do cognitive capabilities of the organizations intervene in the formation* *and success of networks?"*), the model we have built constitutes an implicit answer to them:

we concluded that strategy f) (*cooperation networks*) generates the higher stock of knowledge, in average, among all strategies. It corresponds to a strategy in which networks are formed as a consequence of a collective will and a common goal does not explicitly exist. In addition, it has been observed that most profitable networks have, in general, higher levels of the stock of knowledge (the Pearson correlation coefficient between the profit and stock of knowledge in market X is approximately 0.5);

regarding cognitive capabilities, several mechanisms have been introduced in the model: *learning*, *network support* and *risk*. The effect of learning and network support is significant for the increase of the sock of knowledge; profit is higher in adventurous firms, corresponding to firms with higher attitude to risk.

The research questions (G1 to G10) have been verified during this work. The dynamics of networks observed during the evolutionary process suggest that they are, in fact, dynamic entities running into a perfect adaptation, where the internal synergies and the relationships with other networks shape the course of evolution[149]. Besides, if we consider networks as particular forms of organizations, the empirical evidence shows that certain structural characteristics of an organizational form are surprisingly stable over time (as has been said in Chapter 2,).

Guidelines for future research

Cognitive capabilities

In the future, we shall deeply analyse the effects of cognitive capabilities (individual and social) in the performance of networks. In particular we aim at improving the definition of the *environment understanding* (associated with beliefs, in the architecture

[149] According to Wilkinson et al (2001), *"Firms are operating in complex adaptive systems in which control is distributed through the system. No actor or entity coordinates or directs the behaviour of the network. Firms jointly create both their destiny an the destiny of others; they come to see themselves as parts of business ecosystems in which cooperative and competitive processes act to shape the dynamics and evolution of ecosystem*".

BDI). Furthermore, we plan to give networks the ability to choose collaboration strategies instead of imposing them externally.

Concerning collective intelligence (eventually increased by collective learning) we aim at investigating the mechanism of "energy" that can be measured in coalitions. We follow the suggestion of as Axelrod and Benett (1997) that asks if it is possible to say that some networks are more "intelligent" than others. According to this author, firms should analyse some overall measures in networks (as energy or profit) before choosing (and joining) a particular network. But is profit enough?

Density dependence

Bearing on the key question of how population density affects the intensity of competition, we aim at comparing the models of Barron (2001) (see Equations 2-12, 2-13, 2-14) with the model developed in Chapter 5, using data from Portuguese reality.

Legitimation

We intend to analyse the Legitimation of new forms of business as, for example, Collaborative Virtual Organizations (as web portals, etc)

Knowledge diffusion

The effects of the knowledge diffusion should be analysed in what concerns the underlying network topology. We will try to study the effect of knowledge diffusion in structural holes and in core and peripheral firms. The work of Gulati et al. (2000), and, more recently, Cointet and Roth (2007) are suggestive in this domain.

8. APPENDIXES

8.1. APPENDIX I: Some Definitions of Network Structure

Some researchers use a particular representation to define network structure. In a fundamental paper that aimed at studying the stability and efficiency of social and economic networks, Jackson and Wollinsky (1996) developed a framework that established a paradigm for network structure. This framework was then adopted by many other authors as Cowan *et al.* (2004), Carayol and Roux (2003), Purchase and Olaru (2003) and it is, nowadays, a reference to those researchers who study and explore network structures. Therefore, we will introduce some definitions of the *network structure* as they have been provided by the works of Jackson and Wollinsky (1996) and Jackson (2003). This set of definitions includes Agents, Graphs and Paths.

Agents

Let $\mathcal{N}=\{1, \ldots, N\}$ be a set of *agents* (may be called *individuals*, *organizations* – Jackson and Wollinsky (1996) and Jackson (2003) called them *players*-) who are connected in some network relationship. The network relations among those agents are formally represented by graphs whose nodes or vertices are identified with the agents and whose arcs capture the pairwise relations.

Graphs and links

The complete graph, denoted g^N, is the set of all subsets of $\mathcal{N}$ of size 2. The set of all possible graphs of $\mathcal{N}$ is then $\{g|g \subset g^N\}$. Let ij denote the subset of $\mathcal{N}$ containing i and j and is referred to as the *link* ij. The interpretation is that if $ij \in g$, then nodes i and j are directly connected (or adjacent), while if $ij \notin g$, then nodes i and j are not directly connected. It is important to note that within this context graphs are non-directed, meaning that it is not possible for one agent to link to another without having the second individual linked to the first. Most economic applications consider this situation as mutual consent and as such non-directed graphs will be our central focus.

Let g+ij denote the graph obtained by adding link ij to the existing graph and g-ij denote the graph obtained by deleting link ij from the existing graph g, i.e.:

$$g+ij = g\cup\{ij\}$$ **Equation 0-1**

$$g-ij = g\backslash\{ij\}$$

Paths and components

Let $N(g) = \{i|\exists j : ij\in g\}$, i.e., the set of all links that involve agent i A *path* in g connecting i_1 and i_n is a set of distinct nodes $\{i_1, i_2,\ldots, i_n\}\subset N(g)$ such that $\{i_1i_2, i_2i_3,\ldots, i_{n-1}i_n\}\subset N(g)$. A *path* is therefore a contiguous way starting with the agent i_1 and ending in agent i_n .

Paths divide a network into different connected subgraphs. These subgraphs are commonly referred to as *components.* Consequently the graph $g' \subset g$ is a component of g, if for all $i\in N(g')$ and $j\in N(g')$, with $i\neq j$, there exists a path in g' connecting i and j, and $ij\in g$ implies that $ij\in g'$

Values, efficiency and allocation rules

The network structure is the key determinant of the level of productivity or utility to the society of agents involved. There are methods that keep track of the overall value generated by a particular network. A value function is a function $v: G\rightarrow\Re$. Specifically, the value of a graph is represented by $v: \{g|g\subset g^N\}\rightarrow\Re$. The set of all such functions is denoted $\mathcal{V}$. In some applications, the value v will be an aggregate of individual utilities or productions, $v(g)=\Sigma_i u_i(g)$, where u_i represents the utility of the agent (or node) i, such that: $u_i: \{g|g\subset g^N\}\rightarrow\Re$.

A graph $g\subset g^N$ is strongly efficient if $v(g)\geq v(g')$ for all $g'\subset g^N$. The term *strong efficiency* indicates maximal total value of a network when compared to another. Allocation rules, Y, describe how the value associated with each network is distributed to individual agents. Therefore, $Y_i(g,v)$ is the payoff to agent i from graph g under the value function v. An allocation rule may me formulated as

$Y: \{g|g \subset g^N\} \times V \rightarrow \Re^N$ such that $\Sigma_i Y_i(g,v)=v(g)$ for all v and g.

Stability

As the interest is in understanding which networks are likely to emerge in various contexts, we need to define a notion which captures the *stability* of a network. The definition of a stable graph represents the idea that agents have discretion to form or to dissolve links. The formation of a link requires the consent of both parts involved, but disconnection can be done unilaterally. Jackson and Wollinsky (1996) consider that a graph is *pairwise stable* in respect to value v and allocation rule Y if:

(i) for all $ij \in g$, $Y_i(g,v) \geq Y_i(g\text{-}ij,v)$ and $Y_j(g,v) \geq Y_j(g\text{-}ij,v)$

(ii) for all $ij \notin g$, if $Y_i(g+ij,v) > Y_i(g,v)$ then $Y_j(g+ij,v) < Y_j(g,v)$

In proposition (i), no agent wishes to delete a link that he or she is involved in. Besides, this proposition shows that for both agents i and j, the payoff obtained when the link exists is higher than the one observed if they delete it. Proposition (ii) requires that if some link is not in the network and one of the involved agents would benefit from adding it, then it must be because the other agent may suffer from the addition of the link..

8.2. APPENDIX II: Representation of Actions in MAS: a Framework

Ferber (1999) presented a framework that is oriented towards representing the actions of agents in a simulated world. Although it has not been used in the present our work, the framework is practical for modelling Multi-Agent Systems.

Within this framework the *states* and the *transformation of states* are defined in a fist step: a *state* σ_s of the environment is a set of atomic facts that describe the environment at a particular moment. In other words, the term *state* is used to describe any overall configuration of a system. States can be transformed into new states by *actions.*

The following definitions help to formalize this:

- The set of all possible states of the environment E, is defined by

  $\Sigma=\{\sigma_1, \sigma_2, \ldots, \sigma_\Omega,\}.$ **Equation 8.2-1**

- An action is defined as something that produces a new state. Examples of actions are, for instance, *go* (to relocate an agent), *join* (to associate agents in networks), *move* (to move objects), etc;
- Operators are functions of the type $Op \times \Sigma \rightarrow \Sigma$. There are important functions as Execute (*Exec*) or React (*React*) which after being combined to an appropriate action and state, σ_j , may result (or not) in a new state σ_k. For example, in the syntax *Exec(move(A, C, +1),* σ_1*),* the operator Exec executes the action of moving an object C by agent A one step forward (+1) starting from situation σ_1. In this case, one action (*move*) has been combined with a state (σ_1) and executed by an operator (*Exec*).
- If more than one action is combined with states, then it may give back *influence, i.e:*

  Exec: actions $\mathbf{x}\ \Sigma \rightarrow \Gamma$

where Γ is the set of influences, γ, resulting from the simultaneity of actions completed in a given state.

- The environment can *react* to influences. *React* is the function that specifies how this reaction is made:

$$React: \Sigma \times \Gamma \rightarrow \Sigma$$

Reaction is a function that operates in the set of *states* combined with the set of *influences* and returns a final state. Reaction is the response of the environment to the influences. That is what happens, for example, when two actions are incompatible: one object, *o*, located in L_1 cannot move to two different places L_2 and L_3 at the same time. So, the reaction of the movements (the same object *o* moving in two opposite directions with the same speed) produces a reaction of the environment that forces the object *o* to return to the previous state $\sigma \in \Sigma$, where σ corresponds to the situation where object *o* is in the location L_1.

- An action can be applied to a local modification instead of the whole environment. In such a situation, the environment is represented only by a partial set of states confined to a particular space (as the neighbourhood in *cellular automata* applications). The cell of an automaton is represented by a transition function. It follows an example concerning this situation.

Example 1

Let us imagine that an agent A was situated in location L_1 and that this fact corresponds to state σ_1. Thus, the following example describes the function (*Exec*) that executes an action (*go*) that consists in producing a new state, σ_2 as a result of the displacement of the agent A from the location L_1 to the location L_2 (both locations situated in the environment E).

$$\sigma_2 = Exec\ (go\ (A, L_1, L_2), \sigma_1)$$

In this example, the *action* is *go* (A, L_1, L_2) and the function *Exec* executes the action.

The next example describes the situation where action is viewed as a result of influence. It could be the case of life and death in a *cellular automata* simulation. The Game of

Life created by John Conway (Gardner, 1970) is a famous example of application of cellular automata, where the *cells* or *automata* "live" in a grid.

Example 2

In the Game of Life, as explained above, a cell can only survive if there are three other living cells, for example, in its immediate neighbourhood (that is, the eight cells surrounding it). If these conditions are not satisfied it will die either from the effect of overcrowding (if it has too many living neighbours) or from loneliness (if it has too few). Having *n* agents in this game, (A_1, A_2, …, A_n), and focusing on the survival of a particular cell, A_j, we can therefore define the influence occurring in this game as a set of actions: there is an action corresponding to the evaluation of the state of the cell (death or alive) for all the cells in the neighbourhood of cell A_j (eight cells only).

Simultaneous actions are taken into account. Therefore, the influence γ depends on the overall evaluation of the states of the cells and may result either in the death or in the survival of the cell A_j: consequently the set of possible influences is *to live* (γ_1) and *to die* (γ_2). Considering that the cell only lives if there are three living cells in its neighbourhood, then this can be formalized as follows:

$$\gamma \in \{ \gamma_1, \gamma_2 \} = \begin{cases} \gamma_1 = \{\textit{survive}\ (A_j)\}; & \text{if } \varepsilon=3 \\ \gamma_2 = \{\textit{die}\ (A_j)\}; & \text{if } \varepsilon \neq 3 \end{cases}$$

8.3. APPENDIX III:

Example of R function: *Firmdeath()*

This function is used in NetOrg to determine if a firm is going to fail. The function is based on the density dependence model and considers the piece-wise linear functions of Section 5.3 for the density survival interval, size and age. The function is invoked in the middle of the algorithm. At the end of the code, variable *FirmStatus* returns the value 1 if the firm stays alive and 0 otherwise.

```
firmdeath<-function()   {
  ImitationFlag<-0
  A<-runif(1,0,1)
  B<-runif(1,0,1)
  C<-runif(1,0,1)
  D<-runif(1,0,1)
  TemporalInfluence<-0.2 # To model firm deaths according to reality
  #compute the equation line for the probability p1 of surival of
'ponto'.
######### probability for Density Survival
  ponto1<-SumNeibrs[j]
  x1<-E1
  y1<-0
  x2<-(E1+E2)/2
  y2<-1
  x3<-E2
  y3<-0
  if (ponto1 <E1 | ponto1>E2) p1<-0 else p1<-1
  if (ponto1 >= E1 & ponto1 <= x2) p1<-((y2-y1)/(x2-x1))*(ponto1-
x1)+y1
  if (ponto1 > x2 & ponto1 <=E2) p1<-((y3-y2)/(x3-x2))*(ponto1-x2)+y2
  ######### probability for Size
  s1<-0
  v1<-0
  s2<-SizeLimit
  v2<- ProbabilityforSizeLimit
  ponto2<-Pi[geracao,j]
  if (ponto2 > s2) p2<-ProbabilityforSizeLimit
  if (ponto2 <= s2) p2<-((v2-v1)/(s2-s1))*(ponto2-s1)+v1
  ########## probability for Age
  t1<-0
  u1<-0
  t2<-AgeLimit
  u2<- ProbabilityforAgeLimit
  ponto3<-Age[j]
  if (ponto3 > t2) p3<-ProbabilityforAgeLimit
  if (ponto3 <= t2) p3<-((u2-u1)/(t2-t1))*(ponto3-t1)+u1
  if (B>p1 & (C>p2 | D>p3)) FirmStatus<-0 else FirmStatus<-1
 FirmStatus
}
```

8.4. APPENDIX IV: Collaboration networks: analysis of some case studies (data and trends)

This appendix contains details concerning data of collaboration networks. Case studies are organized in three different areas: Automobile Manufacturing, Textile industry and e-Marketplaces.

Automobile Manufacturing

In many countries, the automotive sector has long been one of the main parts of the metalworking industry. Within the reasons that have motivated this fact, we could include the following[150]:

(1) this sector includes the production of both vehicles and components;

(2) it employs a significant percentage of people in industrialized countries;

(3) it is viewed as a traditional stronghold of trade unions and collective bargaining;

(4) it is viewed as a major source of innovation in areas such as the organisation of production and industrial relations, including the creation of networks.

Once characterised by the prevalence of large national firms, which often evolved into multinationals, the sector has been undergoing significant restructuring in recent decades. Due to company reorganisation initiatives and to the introduction of new technologies and organisational models, the sector experienced an increase in mergers and acquisitions, and the establishment of partnership agreements between automotive firms.

These and other trends from the Automobile manufacturing sector have been observed and synthesized in the following paragraphs. The first set of evidences concerns some

[150] See EIRO – The European Industrial Relations Observatory on-line in http://www.eiro.eurofound.eu.int/2003/12/study/

trends in the automobile industry, while the following concern the facts related to the **challenges**, the **survival** of firms and the type of **relationships** between buyers and suppliers. The work of Swaminathan et al. (2002) is of great importance in terms of providing empirical evidences on automobile manufacturing.

Trends

Following the report from *EMCC - European Monitoring Center on Change* (2004), as car makers seek to cut costs, they outsource more and more to the supply industry. This externalizes a proportion of fixed (overheads) and variable (materials) costs, and shares the risks for new developments. Outsourcing also allows greater economies of specialization and scales, since suppliers are more experienced in certain functions and can supply several carmakers.

Some manufacturers have sold off their in-house component companies in order to concentrate resources and raise funds (General Motors and Ford's two component arms became Delphi and Visteon, respectively). (EMCC, 2004)

Most important companies, that serve as buyers, as Ford, General Motors, VW, etc, known as OEM (Original Equipment manufacturers) concentrate, nowadays, their efforts in the design and assembly phases and transferred other responsibilities to the suppliers, promoting the emergence of "global" suppliers (Selada, et al. 1999)

OEMs tend to migrate to emergent markets of Asia and South America as a way to rationalize production, need for dislocation and capacity reduction (Selada et al. 1998). The reduction of some OEM suppliers has always been observed because of their involvement in the conception of more complex modules and systems and because they follow their better clients in the globalization (Selada et al. 1998)

There has been a kind of competition between nations and regions in the pursuit of the creation of local advantages to attract greater clients (manufacturers) and suppliers. These local advantages are not based in cost reduction anymore, but in intangible factors and intensive knowledge (Selada et al. 1998)

Survival

We have also found some facts, concerning the **survival** of organizations within the context of organizational networks in the automobile sector. Structural changes that took place in the auto-parts industry involved a drastic decrease in the number of automotive suppliers, large requirements in terms of economic and financial capacity and risk and cost sharing with constructors of components development

We expect supplier firms that have long-term relationships with their buyers will enjoy enhanced survival prospects - this is actually one of the Hypothesis of Swaminathan *et al* (2002). Therefore, we may say that *stable and long term links improves supplier firm's probability of survival.*

The higher the current and potential autonomy of a supplier firm, the less likely the supplier will fail. This potential autonomy will have, however, greater influence on the failure rates of suppliers of modular components than of suppliers of architecture components.

Swaminathan *et al* (2002) have also shown that status presents more influence on the failure rate of suppliers of architectural components than on the failure rate of suppliers of modular components.

Industrial relationships

Ritter (1999) study the predisposition and competence to develop relationships and networks. While the vertical networks respect to relationships that a firm maintains with other located in a different area, the horizontal networks refer to relationships of firms located in the same region. (Dimara *et al*. 2003). Consequently, we could say that if the distance between two firms is short, they have more probability to cooperate horizontally; on the other hand, if the distance between two firms is high, they have more probability to cooperate vertically.

Constructors promoted both a concentration process of component suppliers and room for creating some important collaborations and alliances among suppliers and suppliers-constructors. Long-term supply relationships tend to be superior to short term relationships when products are complex, the technology is changing, and there are complicated interactions among components; also, if information transfer is difficult and uncertain or when a trading relationship requires specialized human skills.

The benefits of long term supply relationships arise from three related sources: The development of knowledge of each partner, the development of trust and the development of relationship-specific routines (Swaminathan *et al* 2002).

Long-term relationships generate performance advantages for their members; Dyer and Nobeoka (2000) found a positive relationship between supplier-automaker specialization and performance.

The benefits of long term relationships will be greater for suppliers of architectural components than for suppliers of modular components. Suppliers that become dependent on their buyers often face performance problems resulting from opportunistic behaviour by partners or from lack of access to necessary information.

Due to the difficult in testing all the evidences that are available from previous studies in the Automobile Industry, we will focus on some of these facts, and will resume them as follows:

- Reduction of some OEM suppliers;
- Migration of OEM to emergent markets such as Asia and South America, as a way to rationalize production;
- Contemporaneous density has a negative impact on the mortality of organizations (Carrol and Hannan, 1989).
- Density at founding has a positive impact on the mortality of organizations (Carrol and Hannan, 1989);
- Networks are formed around OEM organizations (several topologies are observed);
- If the distance between two firms is short, they have more probability to cooperate horizontally; If the distance between two firms is high, they have more probability to cooperate vertically;
- Stable and long term links improve supplier firms' probability of survival.

Textile Industry

We have consulted some information concerning the Textile Industry, and found some evidences, most of them brought up by CENESTAP, the Portuguese Center for Applied Textile Studies. We gathered some case studies from the Portuguese, Spanish, and Italian markets that have been made in the context of "Rede Têxtil" (CENESTAP, 2000a, to 2000e, 2001), a project supported by the Portuguese Government in order to improve and promote inter-firm cooperation in Portugal.

In the following, we present a summary of the results obtained in those case studies that can bring about important details about cooperation and the formation of networks.

***Evidences concerning past information*:**

In almost all of the studies from CENESTAP there were between six and thirty firms, directly involved in the creation of networks. One or more new firms were created to formalize the network. In some cases, these new firms were the "image" and the "brand" of the networks.

Evidences concerning collaborative goals of the networks:

We have observed the integration of several (or all) the phases of the production chain in one same formal network. Some of the networks were created as a coalition of small and medium companies from the textile sector in order to distribute the orders. With this kind of cooperation firms, the time to deliver an order will be minimized, and the production capacity of the network will be increased;

New firms that were created (and constitute the image of the network) are often viewed as an interface for clients and suppliers in the production chain.

In terms of horizontal cooperation, those new firms sometimes are responsible for R&D, commercialization, and/or quality issues.

Evidences concerning the evolution of the network

In almost every situation, the responsible for each case-study confirms that the network has explored the complementarity associated with the firms' specificities at several levels; Cooperation had a positive impact on sales, in product upgrading (innovation) and in the concentration of tertiary functions as Marketing, Quality control, R&D, etc.

Cooperation has endorsed the development of coordination mechanisms to reduce the time for order delivery. Some networks have gained the power of negotiation next to their suppliers and some cooperative firms have increased their profits because of the reduction of costs due to the rationalization of some of their functions.
Consequently, some evidences were emerged:

- Small firms can therefore observe new business opportunities;
- International contacts with foreign markets have increased;
- Profitable networks indicate that one of the reasons for their success is a clear definition of the services to develop in cooperation: a common structure with juridical autonomy;
- New functions are available for some firms (as publicity and advertising) because they are accessible to all the firms in the network;
- In some cases there are signs of rivalry, when the distribution of the orders through the network is not compatible with the goals of some particular firms. There are situations of opportunistic behaviours and asymmetric benefits.

e-Marketplaces

Definitions and evolution

E-Marketplaces, a particular type of e-business, is a term often used to identify businesses that are made through electronic means, generally the Internet. The concept of e-business is associated with electronic commerce (e-commerce). In practice, both terms are often used interchangeably. In the last years, the growth of Internet services improved the way economic transactions were made. Electronic inter-organizational information systems are services that enhanced the way how buyers and sellers exchanged information about prices and product offerings (Oppel *et al.* 2001).

Within the Business-to-business (B2B) exchanges, a new concept has emerged: the B2B E-marketplaces. Generically, an e-marketplace (EM) is an open electronic platform facilitating activities related to transactions and interactions between organizations (Wang and Archer, 2004) including all Internet-based technical solutions that aim at facilitating the establishment of new trading relationships between companies or at supporting existing relationships (eMarket services, EU).

B2B e-marketplaces are information technology systems that bring together several business buyers and several business suppliers and facilitate the transaction process by using the Internet (Oppel *et al.* 2001), and offering the advantage of lower transaction and searching costs (Osterle *et al.* 2001). The portal *eMarket Services* (a service created to favour the development of e-marketplaces[151]), uses the following characteristics of a B2B e-Marketplace:

- it is open to several buyers and several sellers;
- it is a trading platform; the e-marketplace itself does not sell nor buy goods or services traded on the platform;
- it has at least one trading function .

Organizations can certainly derive benefits for participating in electronic markets. Although some e-Markets are actually demising, it seems that e-business and, particularly, e-Marketplaces are consolidating. Besides, it is even questionable whether organizations which do not participate in electronic markets will survive (Osterle *et al*, 2001). According to the research, we are now in the third generation of the e-marketplace concept[152], where collaboration plays an important role.

---

[151] eMarket Services is neither a e-marketplace nor a for profit project. It is a web-based service funded by the trade promotion organizations of Australia, Ireland, Norway, Portugal, Spain, Sweden, and The Netherlands. Staff from these organizations work together to produce services and exchange knowledge on electronic marketplaces. Their aim is to provide knowledge and information about e-marketplaces in different industries all over the world.

[152] Following Wang and Archer (2004), the first generation of EMs aimed at creating a more competitive market, eliminating some interactions between buyers, suppliers and between buyers and suppliers. Organizations moved to the second generation looking for more feasibility in the business models. This second generation focused on transaction fulfilment: *placing and tracking orders online*, *contracting logistics electronically,* etc. However, some firms were still not willing to spend enough money to cover

Collaboration in e–marketplaces covers the sharing of risks, information sharing, business complementarities, among other outcomes and driving forces. As we have shown in previous chapters, these are the characteristics of networks, and, more specifically, cooperative networks that are growing within the e-business area.

E-marketplaces and (existing) networks

We need now to distinguish from the two types of collaboration between organizations as it was explained in Chapters 2 and 3. We shall remember that in our topology, the concepts of cooperation and collaboration in networks can be distinguished in this summarized form: cooperation is the lowest level of commitment between two or more organizations, while collaboration offers the highest level of commitment. Sharing of mutual goals exists in collaboration, while in cooperation, either there are different goals, or the mutual goals among organizations are not clear.

Based on this classification, we consider the possibility of organizations to cooperate or collaborate, and this is done in different manners. Inspired by the e-business industry, we can classify the B2B e-marketplaces, according to the product offerings in *vertical marketplaces,* which are industry-specific marketplaces used to trade manufacturing inputs, and *horizontal marketplaces* which are not industry-specific marketplaces, but rather fulfill a specific function in a enterprise used to trade operating inputs (Oppel *et al.* 2001).

The following table shows the classification of e-marketplace's (EM's) business models, according to the types of existing relationships and the type of offering products. It has been inspired by the works of Oppel (2001) and Wang and Archer (2004).

---

these services. Consequently, EMs entered in their third generation, by adding collaboration facilities that are compatible with corporate needs in order to maintain longterm relationships with business partners.

Type of interaction	Type of relationship / products offering	
	Horizontal (Teamwork)	Vertical (supply chain activity)
Cooperation	*Suppliers Catalogues*	*Market Making*
Collaboration	*Pooled purchasing systems*	*Corporate portals*

Examples of B2B e-marketplace's classification of e-marketplace's (EM's) business models, according to the types of existing relationships and the type of offering products

These four EMs business models also differ in the way they use the technologies offered by e-marketplaces. Therefore, Wang and Archer (2004) consider that EMs cannot support all levels of collaboration equally well. The reason for that to happen is related with communication between the participants of the EMs.

The more structured or standardized is the communication, the easier is to codify it into EM software applications. For communication to be easily structured it needs to have patterns and to be relatively stable. In some cases there is even no need to communicate. In horizontal cooperation, for instance, the needs of communication are low, and EMs almost eliminate the interaction among participants.

On the other hand, in horizontal and vertical EMs collaboration models, communication is much unstructured. System developers always lack some part of human communication such as trust, facial expression, social relationships, and tacit knowledge that are difficult to communicate electronically (Wang and Archer, 2004). Consequently, EMs are best at supporting functions with structured communication; for collaboration purposes the information must be well structured and therefore the applications for dealing with such level of communication are few or inexistent.

Even though EMs are best suited for cooperation than for collaboration purposes due to these communication limitations, we can find real industry examples of the four types of EM business models.

Supplier Catalogues are online catalogues for selling goods from different brands. They are not industry specific; they fulfil a particular requirement inside organizations. That is the case of Grainger (www.grainger.com), a B2B EM in the area of the Office supply.

Grainger's customers are 1.7 million different businesses and institutions across North America. Grainger works with more than one thousand suppliers to provide customers with access to more than 800,000 products from categories including Adhesives, Sealants and Tape, Electrical, Fasteners and Hardware, Heating, Ventilation and Air Conditioning, Lighting, etc.

Market Making EMs, or simply *electronic Market Makers* are industry specific vertical business models. Covisint (www.covisint.com) is a joint venture of General Motors, Ford, DaimlerChrysler, Nissan, Renault and other companies that bridges gaps created by dissimilar business systems and adapts businesses to work with the myriad of business processes and technologies used by partners. It is an automobile-industry specific business Electronic Marketplace.

Pooled Purchasing Systems are typical from teamwork or horizontal networks in which two or more organizations join together to form a network with a common goal.

Newportex (www.newportex.com) is an example of this type of horizontal collaboration. Supported by the Portuguese Government, *Newportex* aims at responding successfully to the challenges and opportunities created by globalization and the new technologies. The newportex.com is an EM equipped to carry out the mission of pooling efforts in the Portuguese Textile Industry. The aim is to help companies that belong to newportex.com to find customers and to assist them in responding quickly and effectively.

Corporate portals are other kind of collaborative networks in e-business, appropriate to supply-chain networks. *Autoeuropa*, with the *Volkswagen supply.com* e-marketplace, is an example of vertical collaboration EM that was developed and implemented strategically to promote a closer interaction between *VW Autoeuropa* and its suppliers for industrial materials and Services, in almost all phases of the supply chain.

The following table summarizes this classification and presents other examples of existing B2B EM business models.

e-marketplaces business models	Examples of e-marketplaces
Supplier Catalogues *(Horizontal cooperation)*	**Grainger** (www.grainger.com) Grainger is an online catalogue for selling goods from different brands in the office supply industry. Grainger's idea is to offer a one-stop shopping to help customers saving time and money by providing them the right products *"to keep their facilities up and running"*. They do that by avoiding the interactions between buyers and suppliers and between suppliers themselves. Grainger's on line catalogue is a powerful e-commerce–based system.
Market Making *(Vertical cooperation)*	**Covisint** (www.covisint.com) Covisint is a joint venture of General Motors, Ford, DaimlerChrysler, Nissan, Renault and others. Covisint bridges gaps created by dissimilar business systems and adapts businesses to work with the myriad of business processes and technologies used by partners. Covisint reduces partner collaboration costs by providing a unique framework for collaboration called an 'Industry Operating System'. Covisint manages all of the document translations and maps required to conform to the myriad of customers or partners' requirements - all with a single connection point. Covisint is specialized in the Automotive industry and also in the Health services.
	GNX (www.globalnetxchange.com) GNX is an e-business solution and service provider for the global retail industry. GNX solutions help retailers, manufacturers and their trading partners reduce costs and improve efficiency by streamlining and automating sourcing and supply chain processes. With a hosted platform, restricted to registered buyers or sellers, GNX simplifies the implementation process - and makes technology investments go further.
	Avnet Marshall (www.avnet.com) Avnet, Inc. (NYSE: AVT), is one of the world's largest B2B distributors of semiconductors, interconnect, passive and electromechanical components, enterprise network and computer equipment, and embedded subsystems from leading manufacturers. Serving customers in 70 countries, Avnet markets, inventories, and adds value to these products and provides supply chain management and engineering services. Avnet Marshall's costumers (as Nokia) outsource part of their procurement process to Avnet Marshall.
Pooled purchasing systems *(Horizontal collaboration)*	**Newportex** (www.newportex.com) The newportex.com is a web portal equipped to carry out the mission of pooling efforts in the Portuguese Textile Industry. The aim is to help companies that belong to *newportex.com* to find customers and to assist them in responding quickly and effectively. The newportex.com is totally geared to making business deals more simple and speeding their conclusion. It is supported by a textile association (ANIVEC) and by the Portuguese Government and aims to respond successfully to the challenges and opportunities created by globalization and new technologies.
Corporate portals *(Vertical collaboration)*	**Autoeuropa e VW supply.com** The VW Autoeuropa factory, a production unit located in Palmela, Portugal, is the largest foreign investment project ever done in Portugal. Its goal was to produce three "MPVs" (Multi Purpose Vehicles) from three different brands: VW Sharan, SEAT Alhambra and Ford Galaxy. VW Autoeuropa has a highly positive impact on the Portuguese economy, especially in what exports are concerned. VW Autoeuropa is highly oriented to follow the new worldwide procurement tendencies as well as VW Group strategy of supplier partnership using Internet. Therefore, an e-marketplace (www.vwgroupsupply.com) was developed and implemented strategically to promote a closer interaction between VW Autoeuropa and its suppliers for industrial materials and Services. The aim of this facility is to reduce administrative tasks, accelerate processes and improve planning accuracy and transparency in the collaboration with suppliers.

Examples of several e-marketplaces according to their business models.

Evidences from e-Marketplaces

In the previous paragraphs, we have been describing the business models and some examples of e-Marketplaces. We are now going to systematize some of the observed facts from the e-business area, in order to serve as empirical evidences to our work.

Evidences concerning the structure of an EM network

An EM is usually composed of much more organizations than a conventional (not electronic) network. Just to have an idea of the **dimension** of these electronic networks, we should remark that Grainger, a suppliers' catalogue in the office supply industry is a network with more than 1,000 suppliers. The Covisint joint venture of General Motors, Ford, DaimlerChrysler, etc. links more than 30,000 companies and has over 266,000 users in 96 countries.

The **duration** of these networks is, for sure, difficult to measure. We admit that the network may change its shape during its existence, due to the frequent entrance and exit of partners. Nevertheless, as Camarinha-Matos and Afsarmanesh (2004) point out, "*most collaborative forms are based on long-term relationships*".

Compared to the other sectors (Automobile and Textile industries), the geographical **distance** between firms has a different importance in the study of these types of networks. The use of communication and information technologies to support agile collaboration brings this approach to a new level of effectiveness. In effect, collaboration at a global scale is expected to substantially increase, as distance will no longer be a major limiting issue, Camarinha-Matos and Afsarmanesh (2004).

Evidences concerning economical aspects

As previous examples have shown, one of the principal interesting aspects in e-Marketplaces, is the reduction of **transaction and search costs**. As Oppel *et al*. (2001) state, "*the value of using Electronic B2B marketplaces for buyers lies in reducing purchasing costs. Purchasing costs comprise the actual purchasing price plus transaction costs*". Furthermore, search costs are also reduced, because of the features existing in e-Marketplaces. In fact, the electronic search and comparison of products is one of the most important advantages of EMs.

Grainger.com, for instance, shows a table comparing the costs of several similar products selected by the user. Thus, the time spent with the search of the product may be drastically reduced.

Evidences concerning the relationships between firms

As we have seen before, some levels of interaction and communication such as trust, facial expression, social relationships, and tacit knowledge are difficult to transmit electronically (Wang and Archer, 2004). For that reason, cooperative EMs as *Suppliers Catalogues* and *Market Makers* are most suited for e-Marketplaces than collaborative EMs. Cooperative EMs have structured and standardized communication and the levels of communication are kept small and appropriate to codify it into EM software applications.

Summary of the evidences in e-marketplaces

- Horizontal and vertical cooperation deals with large networks;
- The links are based in long relationships;
- The physical distance between the elements of the network has very low importance:
- Transaction and search costs are reduced. The electronic search and comparison of products is one of the most important advantages of EMs;

All the evidences presented before are used later in Chapter 6 to support the hypotheses to be tested in this work. This identification will take into account the general questions proposed earlier in this chapter.

8.5. APPENDIX V: Creating time dependent covariates in the Cox Extended Model (Chapter 6) - *SPSS syntax scripts and results*

SPSS does not perform automatically the Cox regression for more than one time dependent covariate simultaneously. Therefore, we created two scripts in *syntax* mode, a common scripting mode in SPSS used to perform more complex computations in order to proceed with the extended Cox model. Therefore, we have considered two choices among the most usual forms for defining time-dependent covariates: Kleinbaum and Klein, 2005):

- the product between the time and the covariate, TimexCovariate;
- the product between the natural logarithm of time and the coavariate, ln(Time)xCovariate.

Considering the product between the time and the covariate, *Time*x*Covariate,* the following script syntax has been run in SPSS:

```
TIME PROGRAM.
COMPUTE T_mcost = (T_)*mcost .
COMPUTE T_stockX = (T_)*stock1 .
COMPUTE T_stockY2 = (T_)*stock3 .
COMPUTE T_netbirths = (T_)*netbirths.
COMPUTE T_DNodes = (T_)*DNodes .
COXREG
  age  /STATUS=Status(0)
  /CONTRAST (Strat)=Deviation
  /CONTRAST (form)=Deviation
 /METHOD=ENTER profit T_mcost  T_stockX stock2 T_stockY2 T_netbirths
netdeaths BNodes T_DNodes form Strat
  /PLOT SURVIVAL HAZARD
  /SAVE=SURVIVAL HAZARD
  /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) .
```

Considering now the product between the natural log of time and the coavariate, ln(*Time*)x*Covariate,* the only change in the code would be in the following lines:

```
COMPUTE T_mcost = LN(T_)*mcost .
COMPUTE T_stockX = LN(T_)*stock1 .
COMPUTE T_stockY2 = LN(T_)*stock3 .
COMPUTE T_netbirths = LN(T_)*netbirths.
COMPUTE T_DNodes = LN(T_)*DNodes .
```

The last formulation of the time dependent covariate considering the natural logarithm of the time, originated the following results:

Omnibus Tests of Model Coefficients[a,b]

-2 Log Likelihood	Overall (score)			Change From Previous Step			Change From Previous Block		
	Chi-square	df	Sig.	Chi-square	df	Sig.	Chi-square	df	Sig.
3888.067	232.637	20	.000	245.247	20	.000	245.247	20	.000

a. Beginning Block Number 0, initial Log Likelihood function: -2 Log likelihood: 4133.315

b. Beginning Block Number 1. Method = Enter

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)
profit	-.031	.017	3.511	1	.061	.969
T_mcost	9.789	2.251	18.902	1	.000	17827.796
T_stockX	.000	.000	10.718	1	.001	1.000
stock2	-.005	.002	5.171	1	.023	.995
T_stockY2	.000	.000	3.301	1	.069	1.000
T_netbirths	-.074	.114	.420	1	.517	.929
netdeaths	.208	.095	4.808	1	.028	1.231
BNodes	.045	.019	6.010	1	.014	1.046
T_DNodes	-.036	.012	9.074	1	.003	.965
form			48.459	3	.000	
form(1)	.592	.364	2.651	1	.103	1.808
form(2)	-.460	.372	1.531	1	.216	.631
form(3)	-.726	.388	3.508	1	.061	.484
Strat			9.916	8	.271	
Strat(1)	-.113	.202	.312	1	.576	.893
Strat(2)	-.007	.187	.002	1	.969	.993
Strat(3)	-.259	.211	1.503	1	.220	.772
Strat(4)	.408	.194	4.441	1	.035	1.504
Strat(5)	-.073	.184	.158	1	.691	.930
Strat(6)	-.062	.169	.133	1	.715	.940
Strat(7)	.036	.194	.035	1	.852	1.037
Strat(8)	.164	.189	.751	1	.386	1.179

8.6. APPENDIX VI: Evolutionary Data Analysis: Additional results

In this appendix we present some results (obtained with the software SPAD) that led to the conclusions drawn in the evolutionary data analysis presented in Chapter 6. Methods such as STATIS, Principal Component Analysis (PCA) and Multiple Factorial Analysis (MFA) were used. The next table contains the coordinates and the contributions of the time periods (T1 to T5) in the anlaysis of the Inter-struture:

	Coordinates				**Contributions**			
	Factor 1	Factor 2	Factor 3	Factor 4	Factor 1	Factor 2	Factor 3	Factor 4
T1	0.23	0.34	-0.31	0.10	-0.05	0.00	0.03	0.36
T2	0.48	0.67	-0.10	-0.14	0.04	0.00	0.11	0.04
T3	0.09	0.08	0.25	0.09	0.09	0.00	0.00	0.24
T4	0.15	-0.23	0.29	-0.02	-0.11	0.00	0.01	0.31
T5	3.45	-1.85	-0.12	-0.02	0.02	0.00	0.85	0.05

Coordinates and the contributions of the time periods (T1 to T5) in the anlaysis of the Inter-struture

Concerning the analysis of the intra-structure after the Principal Component Analysis, we obtained the correlations between the variables and the axes or factors. We considered the variables at period 1 and used the variables related to periods 2 to 5 as supplementary variables. The first two axes or factors retain 88.14% of the total variance of the data. In the following table, correlations between active variables and factors are presented.

	Factor 1	**Factor 2**	**Factor 3**	**Factor 4**
Profit1	0.86	-0.44	-0.13	-0.24
Stock1	0.87	-0.17	0.45	0.12
MargC1	0.74	0.66	0.05	-0.14
Diam1	0.91	0.04	-0.34	0.23

Correlations between the active variables and the axes or factors after the Principal Component Analysis

With reference to the supplementary variables and factors (identified with the numbers of the periods), the following table presents the correlations with the factors:

	Factor 1	Factor 2	Factor 3	Factor 4
Profit2	0.79	-0.10	-0.02	-0.07
Stock2	0.41	0.03	0.04	0.06
MargC2	0.50	0.45	0.12	-0.05
Diam2	0.87	-0.03	-0.17	0.16
Profit3	0.63	-0.04	0.07	0.02
Stock3	0.29	0.09	-0.07	0.07
MargC3	0.23	0.38	-0.01	0.07
Diam3	0.60	-0.10	-0.01	0.15
Profit4	0.31	-0.28	0.05	-0.05
Stock4	0.09	0.11	0.05	-0.02
MargC4	-0.06	-0.12	0.00	-0.07
Diam4	0.17	-0.33	-0.18	-0.01
Profit5	-0.07	-0.32	0.08	-0.09
Stock5	-0.22	-0.01	0.01	-0.01
MargC5	0.10	0.36	0.02	-0.03
Diam5	0.31	0.13	-0.04	0.11

Correlations between the supplementary variables and the axes or factors after the Principal Component Analysis

After the clustering of the networks, it is possible to identify what are the variables that contribute most for each class. These variables are identified as characteristic variables in the following table. A description of the clustering in four classes, is also provided, through the comparison, for those specific variables, the mean of each class with the general mean.

For each variable, the test-value is obtained through the statistic:

$$\frac{\overline{x}_q - \overline{x}}{s_{\overline{x}_q}}$$

where $\overline{x}$ is the general mean of a particular variable, and $\overline{x}_q$ is the mean of the variable for a particular class q. The form of the standard deviation $s_{\overline{x}_q}$ can be found in Escoufier et Pagés (1998). The statistic presented above follows a Normal distribution.

	Test Value	p-value	Mean		Std. Dev.		Characteristic variables*
			Class	General	Class	General	
Class 1	5.73	0.000	0.04	0.02	0.02	0.02	20.MargC5
18 observations	4.82	0.000	0.10	0.04	0.07	0.06	8.MargC2
	4.66	0.000	0.10	0.05	0.07	0.06	12.MargC3
	2.39	0.008	653.49	390.33	664.17	553.40	7.Stock2
	-2.94	0.002	6.40	11.53	5.80	8.78	14.Profit4
	-3.16	0.001	1.75	3.42	1.93	2.66	17.Diam4
Class 2							
9 observations	6.00	0.000	71.25	15.18	34.78	30.15	3.Stock1
	5.93	0.000	9.10	1.82	5.29	3.96	2.Profit1
	5.33	0.000	8.27	2.61	2.89	3.43	9.Diam2
	5.21	0.000	7.57	1.87	4.17	3.54	5.Diam1
	5.14	0.000	15.92	5.43	2.43	6.58	6.Profit2
	4.97	0.000	21.35	9.29	0.67	7.82	10.Profit3
	4.52	0.000	6.93	3.23	1.30	2.64	13.Diam3
	4.21	0.000	22.97	11.53	3.80	8.78	14.Profit4
	3.83	0.000	12549.79	4335.59	9600.61	6919.12	11.Stock3
	3.38	0.000	970.34	390.33	402.19	553.40	7.Stock2
	2.99	0.001	0.09	0.03	0.07	0.06	4.MargC1
	2.66	0.004	5.61	3.42	2.27	2.66	17.Diam4
Class 3							
7 observations	5.35	0.000	22954.06	5557.07	13361.85	9082.25	19.Stock5
	4.62	0.000	145144.78	35947.65	83800.98	65998.23	15.Stock4
	3.17	0.001	6.43	3.42	2.20	2.66	17.Diam4
	2.86	0.002	20.51	11.53	1.75	8.78	14.Profit4
Class 4							
26 observations	-2.50	0.006	11357.58	35947.65	17098.73	65998.23	15.Stock4
	-2.88	0.002	0.12	1.82	0.59	3.96	2.Profit1
	-3.33	0.000	0.00	0.03	0.00	0.06	4.MargC1
	-3.37	0.000	0.07	15.18	0.33	30.15	3.Stock1
	-3.47	0.000	0.04	1.87	0.19	3.54	5.Diam1
	-3.53	0.000	0.97	1.26	0.47	0.56	21.Diam5
	-3.55	0.000	97.73	390.33	245.33	553.40	7.Stock2
	-3.62	0.000	608.91	4335.59	908.81	6919.12	11.Stock3
	-3.95	0.000	1.56	5.43	4.21	6.58	6.Profit2
	-4.11	0.000	0.01	0.02	0.01	0.02	20.MargC5
	-4.17	0.000	0.48	2.61	1.23	3.43	9.Diam2
	-4.17	0.000	1.58	3.23	1.67	2.64	13.Diam3
	-4.33	0.000	4.25	9.29	5.27	7.82	10.Profit3
	-4.37	0.000	0.01	0.05	0.01	0.06	12.MargC3
	-4.59	0.000	0.00	0.04	0.01	0.06	8.MargC2

Description of the classes in the clustering of networks, based on the variables that contribute most for each class

Note: * characteristic variables are identified by a sequential number and the time period

9. References

Andras, P., J. Lazarus, G. Roberts and S. Lynden (2006), "Uncertainty and Cooperation: Analytical results and Simulated Agent Society", *Journal of Artificial Societies and Social Simulation,* vol. 9, nº 1

Audretsch, D.B. and M.P. Feldman (1996), "R&D spillovers and the Geography of innovation and production", *American Economic Review*, vol. 86, nº 3, pp. 630-640

Axelrod, R. (1977), *The Complexity of Cooperation*, Princeton, Princeton University Press

Axelrod, R. (1984), *The Evolution of Cooperation*, USA, Basic Books

Axelrod, R. and S. Benett (1997), "A landscape theory of Aggregation" in R. Axelrod (Editor.) *The Complexity of Cooperation*, Princeton, Princeton University Press

Baldwin, J. and M. Rafiquzzaman (1995), "Selection versus Evolutionary Adaptation: Learning and Post-Entry Performance", *International Journal of Industrial Organization*, vol. 13, pp. 501-522

Baldwin, J. M. (1896), "A new Factor in Evolution", *The American Naturalist,* Vol. 30, pp. 441-451

Barabasi. A.-L. (2002), *Linked: The New Science of Networks,* Perseus

Barron, D. N. (2001), "Simulating the Dynamics of Organizational Populations: A Comparison of three Models of Organizational Entry, Exit and Growth", in A. Lomi and E. R. Larsen (editors), *Dynamics of Organizations - Computational Models and Organizational Theories*, Chapter 7, Menlo Park, ,AAAI Press / The MIT Press, pp. 210-242

Baum, J. A. C., and S. J. Mezias (1992), "Localized Competition and Organizational Failure in the Manhattan Hotel Industry, 1898-1990", *Administrative Science Quarterly*, Vol. 37, No. 4, pp. 580-604

Baum, J. and J. Singh (1994), "Organizational Niches and the Dynamics of Organizational Mortality", *American Journal of Sociology*, vol. 100, nº 2, pp. 346-380

Baumol, W. J. (2002), *Free Market Innovation Machine: Analyzing the Growth Miracle of Capitalism*, Princeton University Press.

Becker, R. A., and J. M. Chambers (1985) , *Extending the S System*, Pacific Grove, CA, USA, Wadsworth & Brooks/Cole

Belew, R., and M. Mitchell (1996), "Introduction", in R. Belew and M. Mitchell (editors), *Adaptive Individuals in Evolving Populations: models and algorithms,* Chapter 1, Proceedings volume XXVI, Santa Fe Institute in the Sciences of Complexity, Santa Fe, Addison Wesley

Besanko, D. and R. Braeutigam (2005), *Microeconomics*, USA, John Wiley & Sons, 2nd edition

Bond, A. H., and L. Grasser (1988), *Readings in Distributed Artificial Intelligence*, Los Altos, Morgan Kauffmann

Burt, R. S. (1992), *Structural Holes: the social structure of competition,* Cambridge, MA, Harvard University Press.

Camarinha-Matos, L. and H. Afsarmanesh (2004), *Collaborative Networked Organizations - A research Agenda for emerging Business Models*, Springer

Campos, P. and P. Brazdil (2005), "Density dependence and the survival of firms - a multi-agent approach", *Proceedings of the 6th Workshop on Agent-Based Simulation*, Erlangen, Germany

Campos, P., P. Brazdil and P. Brito (2006), "Organizational survival in cooperation networks: the case of automobile manufacturing", in L. Camarinha-Matos, H. Afsarmanesh, and M. Ollus (editors), *Network-Centric Collaboration and Supporting Frameworks*, Springer, pp. 77-84

Carayol, N. and P. Roux (2003), "Self-Organizing Innovation Networks: when do small worlds emerge?", *4th European Meeting on Applied Evolutionary Economics (EMAEE)*, Utrecht

Carley, K. (1998), "Adaptative Organizations and Emergent Forms", *Proceedings of the 1998 IEEE/ISIC/CIRAC/ISAS Joint Conference*, Gaithersburg, MD, USA, pp. 2-3

Carley, K. M. and V. Hill (2001), "Structural Change and Learning Within Organizations", in A. Lomi and E. R. Larsen (editors), *Dynamics of Organizations - Computational Models and Organizational Theories*, Menlo Park, AAAI Press / The MIT Press, pp. 63-92

Carroll, G. and M. Hannan (1989), "Density Delay in the Evolution of Organizational Populations: A model and Five Empirical Tests", *Administrative Science Quarterly*, vol.34, pp. 411-430

Carroll, G. and M. Hannan (2000), "*The Demography of Corporations and Industries*" Princeton, Princeton University Press

Carvalhido, P. C. (2005), *Análise Conjunta de Quadros de Dados: Comparação de Alguns Métodos*, Dissertação de Mestrado, Faculdade de Engenharia da Universidade do Porto

Cassiman, B. and R. Veugelers (2002), "R&D cooperation and spillovers: some empirical evidence from Belgium", *American Economic Review*, vol. 92, nº 4, pp. 1169-1184

CENESTAP (2000a), *Estudos de casos nacionais: TRL –Têxteis em Rede, LDA, Rede Têxtil - Modelos de Cooperação nas Indústrias Têxteis e do Vestuário*, V. N. Famalicão

CENESTAP (2000b), *Estudos de casos internacionais: Gran Tessuto, SCCRL, Rede Têxtil - Modelos de Cooperação nas Indústrias Têxteis e do Vestuário*, V. N. Famalicão

CENESTAP (2000c), *Estudos de casos nacionais: Lanhoso Invest, SGPS, Rede Têxtil - Modelos de Cooperação nas Indústrias Têxteis e do Vestuário*, V. N. Famalicão

CENESTAP (2000d), *Estudos de casos internacionais: HOTEC - Hotel Contract, SA, Rede Têxtil - Modelos de Cooperação nas Indústrias Têxteis e do Vestuário*, V. N. Famalicão

CENESTAP (2000e), *Estudos de casos internacionais: Gruppo Tessile Gulliver, A.R.L., Rede Têxtil - Modelos de Cooperação nas Indústrias Têxteis e do Vestuário*, V. N. Famalicão

CENESTAP (2001), *Estudos de casos internacionais: Sección Textil - Associacion Espanola para la Calidad, Rede Têxtil - Modelos de Cooperação nas Indústrias Têxteis e do Vestuário,* V. N. Famalicão

Cline, R. (2001), "Partnering for Strategic Alliances", *Lodging Hospitality,* vol. 57, n. 9, pp. 42

Coase, R. H. (1937), "The Nature of the Firm", reprinted in Williamson and Winter (editors), *The Nature of the Firm - Origins, Evolution, and Development*, New York, Oxford University Press, pp. 18-33

Cointet, J. P., and C. Roth (2007), "How Realistic Should Knowledge Diffusion Models Be?", *Journal of Artificial Societies and Social Simulation,* vol. 10, no. 3

Conover, W. J. (1999), *Practical Nonparametric Statistics*, 3rd Edition, USA, John Wiley and Sons

Corbara, B., A. Drougoul, D. Fresneau and S. Lalande (1993), "Simulating the sociogenesis process of ant colonies with MANTA", in P. Bourgine and F. Varela (editors), *Towards a Practice of Autonomous Systems II,* MIT Press

Corchado, J., J. Pavón, E. Corchado and L. Castillo (2004), "Development of CBR-BDI Agents: A Tourist Guide Application", *Lecture Notes in Computer Science*, vol. 3155, pp. 547-559

Cortés, J. C. R., and L.B. Sheremetov (2002), "Model of Cooperation in Multi Agent Systems with Fuzzy Coalitions", *Lecture Notes in Computer Science,* 2691, pp. 268-280

Cowan, R., N. Jonard and J. B. Zimmerman, (2004), "Networks as emergent structures from bilateral collaboration", *MERIT - Infonomics Research memorandum series*, 2004-17

Cox, D. R. and D. Oakes (1984), *Analysis of Survival Data*, Cambridge, Chapman and Hall

Cruz, F. C. (1999), "Indústria e Território: as economias externas como factor de competitividade dos sistemas produtivos locais", Tese de Mestrado, Escola de Economia e Gestão da Universidade do Minho, Braga, Portugal

Csardi, G. (2005), *IGraph library – free software*, MTA RMKI, Konkoly-Thege Miklos st. 29-33, Budapest 1121, Hungary

Csermely, P. (2006), *Weak Links: Stabilizers of Complex Systems from Proteins to Social Networks,* Germany, Springer

D'Agata, A. and G. Santangelo (2003) "Cognitive distance, knowledge spillovers and localisation in a duopolistic game", Catania, *mimeo*.

Darwin, C. (1985 [1859]), *The origin of species by means of natural selection; or, the preservation of favoured races in the struggle for life*, New York: Penguin.

Davies, P. and M. Favreault (2003), "Interactions between Social Security Reform and the Supplemental Security Income Program for the Aged", *Working Paper,* The Urban Institute Washington

Dazy, F., J-F. Le Barzic, (1996), *L'Analyse des Données Évolutives - méthodes et applciations*, éd. Technip, 1996

De Woot (1987), "Capacité stratégique et performance économique à long terme des enterprises europeennes", Project (PENELOPE), FAST, Comission of the European Community, Brussels

Dimara, E., A. Goudis, D. Skuras, and K. Tsegenidi (2003), "Business Networks and Performance: a spatial approach", *43th Congress of ERSA*, "Peripheries, Centres and Spatial Development in the New Europe", Finland, 27th-30th August

Dixit, A. K. and J. E. Stiglitz (1977), "Monopolistic competition and optimum product diversity", *American Economic Review*, vol. 67, n. 3, pp. 297-308

Doran, J. E., and N. Gilbert (1994), "Simulating Societies,: an Introduction", in N. Gilbert and J. E. Doran (editors) *Simulating Societies: The Computer Simulation of Social Phenomena*, pp. 1-18. London, UCL Press

Dorigo, M., V. Maniezzo, and A. Colorni (1996), "The Ant System: Optimization by a Colony of Cooperating Agents", *IEEE Transactions on Systems, Man and Cybernetics – Part B*, Vol. 26, No. 1, pp. 1-13

Dumitrescu, D., B. Lazzerini, L. C. Jain, and A. Dumitrescu (2000), *Evolutionary Computation*, USA, CRC Press

Dyer J. H. and K. Nobeoka (2000), "Creating and managing a high-performance knowledge-sharing network: the Toyota case", *Strategic Management Journal*, vol. 21, pp. 345-367

Ebers, M. (1997), "Explaining Inter-Organizational Network Formation", in M. Ebers (editor), *The Formation of Inter-Organizational Networks*, New York, Oxford University Press, pp. 3-40

Egelmore, R., and T. Morgan (1988), *Blackboard Systems*, Addison Wesley

Eiriz, V. (2004), "Dinâmica de relacionamento entre redes interorganizacionais", *Inovação Organizacional*, No. 2, pp. 121-153

EMCC - European Monitoring Centre on Change (2004), *Trends and Drivers of change in the European Automotive Industry: mapping report,* Dublin, European Commission

Epstein, J. (2003), "Growing Adaptative organizations: an agent-based computational approach", *Working paper 03-05-029*, Santa Fe Institute, USA

Epstein, J. and R. Axtell (1996), *Growing Artificial Societies, Social Science from the bottom up*, Washington, Brookings Institution Press, The MIT Press

Escoufier, B. and J. Pagés (1998), *Analyses Factorielles Simples et Multiplies: objectifs, méthodes et interprétation*, Paris, Dunod, 3rd edition

Fagiolo, G., A. Moneta, and P. Windrum, (2005) "Empirical Validation of Agent-Based Models", *ACEPOL 2005 - International Workshop on 'Agent-Based Models for Economic Policy Design'*, Bielefeld, July 2005

Ferber, J. (1999), *Multi-Agent Systems - An Introduction to Distributed Artificial Intelligence*, Harlow, Addison-Wesley Longman

Forrester, J. W. (1971), *World Dynamics,* Cambridge, MA, MIT Press

Freeman, J. (1982), "Organizational life cycles and natural selection processes", in B. Staw and L. Cummings (editors), *Research in Organizational Behavior*, chapter 4, JAI Press Inc, pp. 1-32

Gans, G., and M. Jarke, G. Lakemeyer, D. Schmitz (2004), "Deliberation in a MetaData-based modeling and simulation environment for inter-organizational networks", *Information Systems*, No. 30, pp. 587-607

Gardner, M. (1970), "Mathematical Games: On cellular automata, self-reproduction, the Garden of Eden and the game 'life' ", *Scientific American*, 223, pp. 120-123.

Gilbert, N. (2005). "When does social simulation need cognitive models?" In R. Sun (Editor), *Cognition and multi-agent interaction: From cognitive modelling to social simulation.* Cambridge: Cambridge University Press

Gilbert, N. and K. Troitzsch (1999), *Simulation for the Social Scientist*, Buckingham, Open University Press

Gilbert, N., A. Pyka and P. Ahrweiller (2001), "Innovation Networks - A simulation approach", *Journal of Artificial Societies and Social Simulation*, vol. 4, nº 3.

Goldberg, D. E. (1989), *Genetic Algorithms in Search Optimization and Machine Learning*, Addison-Wesley Publishing, Inc., Reading, Massachusetts

Gordon, R. (1991) 'Innovation, Industrial Networks and High-technology Regions', in R. Camagni (ed.) *Innovation Networks: Spatial Perspectives*, London, Belhaven Press, pp. 174-195

Gulati, R., N. Nohria and A. Zaheer (2000), "Strategic Networks", *Strategic Management Journal*, vol. 21, pp. 203-215

Hakansson, H. (1982), (editor), *International Marketing and Purchasing of Industrial Goods – As interaction approach*, Chichester, Wiley

Hakansson, H. (1987), (editor), *Industrial technological development: a network approach*, London, Croom Helm

Hakansson, H. and I. Snehota, (1995), *Business Networks*, London, Routledge

Hakansson, H., V. Havila and A-C. Pedersen (1999), "Learning in Networks", *Industrial Marketing Management*, vol. 28, pp.443-452

Hannan, M and J. Freeman (1988), "The Ecology of Organizational Mortality: American Labour Unions, 1836-1985", *American Journal of Sociology*, Vol. 94: pp. 25-52

Hannan, M. and G. Carroll (1992), *Dynamics of Organizations - Computational Models and Organizational Theories*, New York, Oxford University Press

Hannan, M. and J. Freeman (1977), "The Population Ecology of Organizations", *American Journal of Sociology*, vol. 82, pp. 929-964

Hannan, M. and J. Freeman (1984), "Structural Inertia and Organizational Change", *American Economic Review*, vol. 49, pp.149-164

Hannan, M. and J. Freeman (1989), *Organizational Ecology*, Harvard, Harvard University Press

Hannan, M., D. Barron and G. Carroll (1991), "On the Interpretation of Density Dependence in Rates of Organizational Mortality: A Reply to Petersen and Koput", *American Sociological Review*, Vol. 56, No. 3 (Jun., 1991), pp. 410-415

Heppner, F., and U. Grenander, (1990) "A stochastic nonlinear model for coordinated bird flocks", In: S. Krasner, Editors, *The Ubiquity of Chaos*, AAAS, Washington (1990), pp. 233–238.

Hinton, G., and S. Nowlan (1987), "How Learning Can Guide Evolution", *Complex Systems*, 1, 495-502.

Hitt, M., R. R. Ireland and R. E. Hoskisson (2005), *Strategic Management: Competitiveness and Globalization (concepts and cases)*, 6th edition, Mason, Thomson, South-Western

Holland, J. (2001 [1975]), *Adaptation in Natural and Artificial Systems - an Introductory Analysis with Applications to Biology, Control and Artificial Intelligence*, The MIT Press, 6th printing

Holland, J. H. and J. Miller (1991), "Artificial Adaptive Agents in Economic Theory", *American Economic Review*, vol 81, AEA papers and proceedings, pp.365-370

Hopcroft, J. E. and J. D. Ullman (1979), *Introduction to Automata Theory, Languages, and Computation*, Reading, Addison-Wesley

Hu, X., Y. Shi, and R. Aberhart (2004), "Recent Advances in Particle Swarm", *In Proc. of the 2004 Congr. on Evolutionary Computation*, vol. 1, pp. 90–97,

Huhns, M., and M. P. Singh (1998), *Readings in Agents*, San Mateo, CA, Morgan Kaufmann

INE - Instituto Nacional de Estatística, (2002, 2003, 2004), *Retrato Territorial de Portugal*, Lisboa, INE

Jackson, M. (2003), "A Survey of Models of Network Formation: Stability and Efficiency", in G. Demange and M. Wooder (Editors), *Group Formation in Economics: Networks, Clubs, and Coalitions*, Cambridge, Cambridge University Press, pp. 11-57

Jackson, M. and A. Wollinsky (1996), "A strategic model of Social and Economic Networks", *Journal of Economic Theory*, vol. 41, pp. 44-74

Jaffe, A. B. (1989), "Real effects of academic research", *American Economic Review*, vol. 79, nº 5, pp. 957-970

Jaffe, A. B., M. Trajtenberg and R. Henderson (1993), "Geographic localization of knowledge spillovers as evidenced by patent citations", *Quarterly Journal of Economics*, vol. 108, nº 3, pp. 577-598

Jarillo, J. C., (1988), "On Strategic Networks", *Strategic Management Journal,* Vol. 9, no. 1, pp. 31-41

Jarillo, J. C., (1993), *Strategic Networks – Creating the Borderless Organization*, Butterworth Heinemann, Oxford

Johanson, J., and L-G Matsson, (1987), "Interorganizational Relations in Industrial Systems: a Network Approach Compared with Transaction-Costs Approach", *Int. Studies of Management and Organization*, Vol. XVII, No. 1, pp. 34-48

Kamann, D. J. F., and D. Strijker (1991), The Network Approach: concepts and Applications, in R. Camagni (ed.) Innovation Networks: Spatial Perspectives, London, Belhaven Press, pp. 145-172

Kennedy J., and R. Eberhart (1995), Particle Swarm Optimization, in *Proc. of the IEEE International Conference on Neural Networks*, Piscataway, NJ, pp. 1942–1948

Keyfitz, N. (1977) *Applied Mathematical Demography,* New York, Wiley

Kleinbaum, D., and M. Klein (2005), *Survival Analysis: a self-learning text*, Second edition, USA, Springer

Kogut, B. (2000), "The Network as knowledge: generative rules and the emergence of structure", *Strategic Management Journal*, vol. 21, pp. 405-425

Koleva, G. (2002), "Comparison between Alliances, Networks and Joint Ventures: what management techniques are in place?", *A Research Note*, Copenhagen Business School

L´Hermier des Plantes, H. (1976), *Structuration des Tableaux à Trois Indices de la Statistique*, Thèse de 3ème cycle, Université de Montpellier

Lamarck, J. B. (1914), "On the Influence of the Environment on the Activities and Habits of Animals, and the Influence of the Activities and Habits of these Living Bodies in Modifying Their Organization and Structure", *Zoological Philosophy*. London, reprinted in R. Belew and M. Mitchell (1986) (editors), *Adaptive Individuals in Evolving Populations: models and algorithms,* Proceedings volume XXVI*,* Santa Fe Institute in the Sciences of Complexity, Santa Fe, Addison Wesley, pp. 39-57

Latora, V., and M. Marchoiri (2003), "Economic Small-World Behavior in Weighted Networks, *The European Physical Journal B,* Volume 32, Issue 2, pp. 249-263

Lavit, C. (1988), *Analyse conjointe de tableaux quantitatifs*, Masson

Lechner, C. and M. Dowling (2003), "Firm Networks: external relationships as sources for the growth and competitiveness of entrepreneurial firms", *Entrepreneurship & Regional Development*, vol. 15, pp.1-26

Leibenstein, H. (1968), "Entrepreneurship and Development", *American Economic Review,* vol. 58

Leskovec, J., J. Kleinberd and C. Faloutsos, (2005), "Graphs over time: densification laws, shrinking diameters and possible explanations", *Conference on Knowledge Discovery in Data (KDD '05)*, Chicago Press, pp. 77-187

Littman, M. L., (1996), "Simulation Combining Evolution and Learning", in R. Belew and M. Mitchell (editors), *Adaptive Individuals in Evolving Populations: models and algorithms,* Proceedings volume XXVI*,* Santa Fe Institute in the Sciences of Complexity, Santa Fe, Addison Wesley, pp. 465-478

Loasby, B. (1994), “Organizational capabilities and Interfirm relations”, *Metroeconomica,* Vol. 45, pp. 248-265

Lomi, A. (1995), “The Population Ecology of Organizational Founding: Location Dependence and Unobserved Heterogeneity”, *Administrative Science Quarterly*, vol. 40, pp. 114-144

Lomi, A. and E. R. Larsen (1998), “Density Delay and Organizational Survival: Computational Models and Empirical Comparisons”, *Computational & Mathematical Organization Theory*, vol. 3, nº 4, pp. 219-247

Lotka, A. J. (1925), *Elements of physical biology*, Baltimore, Williams and Wilkins

Malerba, F., R. Nelson, L. Orsenigo and S. Winter (2001), “History-Friendly Models: an overview of the case of the computer industry”, *Journal of Artificial Societies and Social Simulation*, vol. 4, nº 3

Mata, J. and P. Portugal (1994), “Life Duration of New firms”, *Journal of Industrial Economics*, vol. 42, pp. 227-246

Mata, J., P. Portugal and P. Guimarães (1995), “The Survival of New plants: start-up conditions and post-entry evolution”, *International Journal of Industrial Organization*, vol. 13, pp. 459-481

Mattson, L.-G. (1997), “’Relationship Marketing’ and the Markets-as-Networks Approach: a Comparative Analysis of Two Evolving Streams of Research”, *Journal of Marketing Management*, Vol. 13, No. 5, pp. 447-461

Michalski, R., (1983), “Theory and methodology of inductive learning”, *Artificial Intelligence*, Vol. 20, pp. 111-161

Milgram, S. (1967), “The Small-World Problem”, *Psych. Today*, Vol. 1, pp. 62-67

Mitchel, J. C. (1969), “The Concept and Use of Social Network" in J. C. Mitchell (editor), *Social Networks in Urban Situations*, Manchester, UK, Manchester University Press, pp. 1-50

Mitchell, T. (1997), *Machine Learning*, Singapore, McGraw Hill

Mizruchi, M. and B. Potts (1998), “Centrality and power revisited: actor success in group decision making”, *Social Networks*, vol. 20, pp. 353-387

Monod, J. (1971), *Chance and Necessity: An Essay on the Natural Philosophy of Modern Biology*, New York, Vintage

Morin, E. (1977), *La Méthode (1): La Nature de la Nature*, Seuil

Nelson, R. and S. Winter (1982), *An Evolutionary Theory of Economic Change*, Cambridge, Mass. and London, Belknap Press

Newman, M. E. J. (2000), "Models of the Small World: a review", *Working Paper* 99-12-080, Santa Fe Institute furnish

OCES - Observatório da Ciência e do Ensino Superior (2005), *Inquérito ao Potencial Científico e Tecnológico Nacional (Sector Empresas)* – information requested by the author

Oppel, K, E. Hartmann, M. Lingenfelder and H.-G. Gemünden (2001), "Electronic B2B Marketplaces - Impact on B2B Transactions and Relationships", *Proceedings from the 17th, Annual IMP conference*, 9-11 September, Oslo.

Ormerod, P., H. Johns and L. Smith (2002), "Agent-Based Model of Firm Extinction", *Working Paper*, Volterra Consulting, London

Osterle, H., H. Fleisch and R. Alt (2001), *Business Networking - shaping collaboration between enterprises*, Second edition, Germany, Springer

Pavitt, K., (1984), "Sectoral patterns of technical change: towards a taxonomy and a theory", *Research Policy,* vol. 13, pp. 343-373

Penrose, E. (1952), "Biological Analysis in the Theory of the Firm", *American Economic Review*, Vol. 42, No. 5, pp. 804-819

Penrose, E. (1959), *The Theory of the Growth of the Firm*, Oxford, Basil Blackwell

Poston, D. L., and M. Micklin, (editors), (2005), *Handbook of Population*, Kluwer Academic Press

Pujol, J. M., A. Flache, J. Delgado and R. Sangüesa (2005), "How can social networks ever become complex? Modeling the emergence of complex networks from local social exchanges", *Journal of Artificial Societies and Social Simulation*, vol. 8, nº 4

Purchase, S. and D. Olaru (2003), "Business Networks: simulation analysis", *19th IMP Conference*, Lugano, Switzerland

Quinlan, J.R. (1997), *C5.0 Data Mining Tool*, www.rulequest.com

R Development Core Team (2005), "R: A language and environment for statistical computing", *R Foundation for Statistical Computing*, Vienna, Austria

Rao, A. S. and Georgeff, M. P. (1995), "BD1 Agents: From theory to practice", In *Proceedings of the International Conference on Multi-Agent Systems (1CMAS-95)*, San Francisco, USA, June, 1995

Rao, A. S. and M. P. Georgeff (1991), "Modelling rational agents within a BDI-architecture", In J. Allen, R. Fikes and E. Sandewall, (editors), *Proceedings of the Second International Conference on Principles of Knowledge Representation and Reasoning*, Morgan Kaufmann Publishers, San Mateo, CA, 1991

Ratti, R., (1991), 'Small and Medium-Size Enterprises, Local. Synergies, and Spatial Cycles of Innovation', in R. Camagni (editor), *Innovation Networks: Spatial Perspectives*, London, Belhaven Press, pp.71–88

Reynolds, C. W. (1987), "Flocks, Herds, and Schools: A Distributed Behavioral Model", In Computer Graphics (SIGGRAPH '87 Conference Proceedings), Vol. 21, No. 4, pp. 25-34

Ritter, T. (1999), "The Networking company antecedents for coping with relationships and networks effectively", *Industrial Marketing Management*, vol. 28, pp. 467-479

Rogers, A. and P. von Tessin (2004), "Multi-Objective Calibration For Agent-Based Models", *Proceedings of Agent-Based Simulation 5*, Lisbon, Portugal.

Roseira, C. (2005), *O Impacto da Gestão de Fornecedores no Desempenho das Empresas Industriais*, Tese de Doutoramento em Ciências Empresariais, Faculdade de Economia da Universidade do Porto

Schelling, T. (1978), *Micromotives and Macrobehaviour*, New York, Norton

Schoenfeld, D. (1982), "Partial residuals for the proportional hazards regression model", *Biometrika*, No. 69. pp. 239-241

Schumpeter, J. A. (1996 [1942]), *Capitalism, Socialism and Democracy*, London, Routledge

Scitovsky, T. (1954), "Two concepts of external economies", *Journal of Political Economy*, vol. 62, pp. 142-151

Selada, C., J. R. Felizardo and L. P. Féria (1999), *A integração das Infra-estruturas tecnológicas na Rede de Excelência para o desenvolvimento da Indústria automóvel em Portugal: uma metodologia de avaliação*, Ministério da Economia, DT - 23-99

Selada, C., T., Rolo, J. R. Felizardo and L. P. Féria (1998), *A Dinamização da Cooperação Interempresarial no Sector de Componentes de Automóvel: O caso de Estudo ACECIA, ACE*, Ministério da Economia, DT - 16-98

Silva, S. (2006), *Microfoundations of Industrial Dynamics and Economic Growth*, Doctoral Thesis in Economics, Faculdade de Economia da Universidade do Porto

Simon, H. (1955), "A Behavioral Model of Rational Choice", *The Quarterly Journal of Economics*, vol. 1, pp. 99-118

Smith, A., (1776), *An Inquiry into the Nature and Causes of the Wealth of Nations*, 5a. ed. (edited by Edwin Cannan, 1930),

Solé, F., and J. Valls, (1991), "Networks of Technological Cooperation between SMEs: Strategic and Spatial Aspects", in Camagni (ed.) *Innovation Networks: Spatial Perspectives*, London, Belhaven Press, pp. 174-195

SPSS for Windows (2007), Rel. 15.0, Chicago: SPSS Inc.

Stinchcombe, A. (1965), "Social structure and organizations", in J. March (editor), *The Handbook of Organizations*, Chapter 4, Chicago, Rand McNally & Company, pp. 142-193

Sun, R. (2001), "Cognitive Sciences meets Multi-Agent Systems: a prolegomenon", *Philosophical Psychology*, vol. 14, nº 1, pp. 5-28

Swaminathan, A., G. Hotker and W. Mitchell (2002), "Network Structure and Business Survival: the case of U.S. Automobile Component Suppliers", *Working Paper*, University of Illinois

Tecuci, G. (1998), *Building Intelligent Agents: An Apprenticeship Multistrategy Learning – Theory, Methodology, Tool and Case Studies*, London, Academic Press

Tesfatsion, L. (2006), "Agent-Based Computational Economics: a constructive approach" in L. Tesfatsion and K. Judd (editors), *Handbook of Computational Economics*, vol. 2, Amsterdam, North-Holland, pp. 831-880

Todd, P., (1996), "The Causes and Effects of Evolutionary Simulation in the Behavioural Sciences", in Belew, R., and M. Mitchell (editors), *Adaptive Individuals in Evolving Populations: models and algorithms,* Proceedings volume XXVI, Santa Fe Institute in the Sciences of Complexity, Santa Fe, Addison Wesley, pp. 211-231

Usher, L. and M. Evans (1996), "Life and Death along gasoline alley: Darwinian and Lamarckian processes in a differentiating population", *Academy of Management Journal*, vol. 39, nº 5, pp. 1428-1466

Venkataraman, S. and A. Van de Vem (1998), "Hostile Environmental Jolts, Transaction set and new business", *Journal of Business Venturing*, vol. 13, pp. 231-255

Volterra, V. (1927), "Variations and Fluctuations in the Number of Individuals in Animal Species Living Together" in R. N. Chapman (1931), *Animal Ecology*, McGraw Hill

Wang, S. and N. Archer (2004), "Supporting Collaboration in Business-to-Business Electronic Marketplaces", *Information Systems and e-Business Management*, 2(2) pp. 271-288

Watts, D. J., and S. H. Strogatz, (1998), "Collective Dynamics of `Small-World' Networks," *Nature*, vol. 393, no. 6684, 1998, pp. 440-442

Wersching, K. (2005), "Agglomeration in an Innovative and Differentiated Industry with Heterogeneous Knowledge Spillovers", *Workshop Regional Agglomeration, growth and multilevel governance: the EU in a comparative perspective*", Ghent (Belgium), 24-25 Nov.

Wilhite, A., (2006), "Economic Activity on Fixed Networks" in L. Tesfatsion and K. Judd (editors), *Handbook of Computational Economics*, vol. 2, Amsterdam, North-Holland, pp. 1015-1045

Wilkinson, I. F., (2006), "The Evolvability of Business and the Role of Antitrust", *Antitrust Bulletin*, Vol. 51, No. 1, pp. 111-142

Wilkinson, I. F., J., Wiley, and A. Lin (2001), "Modelling the Structural Dynamics of Industrial Networks", *InterJournal Complex Systems*, volume Article #409.

Williamson, O. (1975), *Markets as Hierarchies: Analysis and Antitrust implications*, New York, The Free Press

Williamson, O. (1985), *The Economic Institutions of Capitalism*, New York, The Free Press

Willimanson, O. (1991), "The Logic of Economic Organization", in O. Williamson and S. Winter, *The Nature of the Firm - Origins, Evolution, and Development*, New York, Oxford University Press, pp. 90-116

Windrum, P., G. Fagiolo and A. Moneta (2007), "Empirical Validation of Agent-Based Models: Alternatives and Prospects", *Journal of Artificial Societies and Social Simulation,* vol. 10, no. 2,

Winer M. and K. Ray (1994) *Collaboration Handbook: Creating, Sustaining and Enjoying the Journey*, Saint Paul MN, Amherst H. Wilder Foundation

Wissen, L. (2000), "A micro-simulation model of firms: applications of concepts of the demography of the firm", *Papers in Regional Science*, vol. 79, pp. 111-134

Wissen, L. (2004), "A spatial interpretation of the density dependence model in industrial Demography", *Small Business Economics*, vol. 22, pp. 253-264

Wooldridge, M. (2002), *An Introduction to MultiAgent Systems*, Sussex, John Wiley and Sons

Young, H. P. (1998), Individual Strategy and Social Structure - An evolutionary theory of Institutions, Princeton, Princeton University Press

Young, L., and Wilkinson, I. F., (1989) "The Role of Trust and Co-operation in Marketing Channels: A Preliminary Study", *European Journal of Marketing*, 1989, 23, 2: 109-122

Zhang, J. (2003), "Growing Silicon Valley on a Landscape: an agent-based approach to high-tech industrial clusters", *Journal of Evolutionary Economics*, vol. 13, pp. 529-548