The evolution of knowledge systems

François Lafond

The evolution of knowledge systems

DISSERTATION

to obtain the degree of Doctor at
Maastricht University,
on the authority of the Rector Magnificus, Prof. dr. L.L.G. Soete,
in accordance with the decision of the Board of Deans,
to be defended in public on Friday 17 October 2014, at 12.00 hours

by

François Lafond

Supervisor
Prof. Dr. Robin Cowan

Assessment Committee

Prof. Dr. Théophile Azomahou (Chair)
Prof. Dr. Jean-Jacques Herings
Prof. Dr. Koen Frenken (Utrecht University)
Dr. Matteo Marsili (Abdus Salam International Centre for Theoretical Physics, Italy)

Summary

The creation and allocation of economic wealth ultimately relies on the creation and diffusion of knowledge. As a result, understanding the dynamics, organization and viability of economies requires in-depth analysis of knowledge systems. This thesis proposes to study knowledge systems as self-organizing two-mode networks. Two-mode networks have two types of nodes, and the links are only between nodes of different types. These are self-organizing in the sense that simple rules of evolution lead to a rich but patterned dynamics. The thesis builds on the literature on social (agent-agent) and epistemic (idea-idea) networks to study socio-epistemic co-evolution (agent-idea networks).

It is found that: (i) stable power law distributions of ideas' popularity naturally emerge from innovation and face-to-face diffusion; (ii) this dynamic is compatible with the observed (shifted) power law distribution of citations, and (iii) the generalized beta size-rank relation observed for patent classes can be explained by a slowdown in the growth of the number of classes. A general lesson from this work is that knowledge systems often exhibit non-equilibrium and non-linear dynamics, which may cast doubts on their long term viability.

The thesis starts with a general introduction (chapter 1) followed by a review of the literature on knowledge and development (chapter 2) explaining why the study of self-organizing knowledge systems should be a core topic of economics. Next, an overview of growing one-mode and two-mode network models is given (chapter 3), together with some original results. The three main chapters follow.

Chapter 4 presents a model of creation and diffusion of ideas in a social network. Agents learn random ideas of random friends, creating a self-reinforcing dynamic in ideas' diffusion. However, this exponential diffusion is constrained since population is bounded, leading to a logistic diffusion curve. At the steady-state, the distribution of ideas' popularity (the number of agents knowing an idea) is a specific Generalized Hypergeometric Distribution, which tends to the well-known Yule-Simon distribution as the population size goes to infinity. A mean-field self-consistency equation for the partition factor of the attachment kernel highlights that a stable self-organization takes place because the "overlap" among agents' ideas sets,

determining and determined by which ideas diffuse, reaches a stable state.

Chapter 5 uses the results of chapter 4 to explain the structure of citation networks. In chapter 5, it is assumed that, when new ideas (papers) appear, they cite random ideas previously known by the inventor. In contrast to existing models, which explain the exponent of power law citation distributions by the relative prevalence of bibliographic and random search (reading papers cited by other papers, or found at random), the model shows that limited attention and face-to-face diffusion can also explain the observed patterns.

Chapter 6 studies the size of patent categories. The size-distribution of US patent *sub*classes is well fitted by a (shifted) power law. However, at the level of patent classes, Zipf's law is clearly violated, and the size-rank relationship is well fitted by a generalized beta distribution. To explain this pattern, a nonlinear modification of the Yule-Simon-Naranan principle is proposed. While according to the latter, both individual categories and the number of categories grow exponentially, here it is assumed that the number of categories follows an asymmetric logistic (Richards) curve.

Acknowledgements

It is my great pleasure to be able to thank Prof. Robin Cowan for providing continued support and intellectual freedom. I am also grateful to Théophile Azomahou, Koen Frenken, Jean-Jacques Herings and Matteo Marsili for having accepted to serve as committee members. I do understand how dry is the thesis, and I appreciated the timely comments on it.

At an early stage, I benefited from the comments of various UNU-MERIT researchers, and of DIMETIC participants, where my discussants Koen Frenken and Luis Izquierdo have challenged me in a decisive way. The thesis has benefited from comments at various audiences, in Maastricht, Utrecht, Eindhoven, Hamburg, Reijkjavik, Bordeaux, Oxford, Hangzhou, and Santa-Fe. At these occasions I have had the chance to engage in productive intellectual exchanges with many people, but I would like to acknowledge particularly Luis and Paolo. I also benefited from the comments of my new colleagues, Debbie, Doyne, HyeJin, Ioannis and Jose.

I would also like to thank UNU-MERIT staff members, who have been as kind and helpful as one can be, particularly those I interacted the most with, Ad, Eric, Eveline, Herman, Mourik-Jan, and Wilma.

I have been lucky to be part of a great cohort at UNU-MERIT: Ale, Daniel, Giorgio, Josy, Samy, Tatevick. Thanks to them I survived and eventually enjoyed bike rides on the ice, two-variables difference equations, kaassoufle-scrambled eggs, non-english speaking movies with dutch subtitles, etc. For a variety of reasons I would like to thank Ajay, Andy, Andrea, Andrés, Biljana, Boris, Bulat, Charlotte, Daniel V, Eduardo, Francesca, Hibret, Iman, Jennifer, Jojo, Jun, Lilia, Michiko, Marco, Michael, Shuan, Simone, Omar, Rodolfo, Sachin, and Zakaria.

I am indebted to Stefi for, *"inter alia"*, her help, patience, and support. Since I would hardly find the words to thank my family, let this book be dedicated to them.

Finally, I am grateful to Derek Lilly for letting me use a photo he took as the cover of this book.

Contents

Chapter 1

Introduction

By its very definition, economics studies how people organize to create and allocate wealth. But if wealth ultimately relies on knowledge, then economics should study the dynamics and organization of knowledge systems. In fact, the creation of value relies on what people do, individually and collectively. What people do and how they do it – the routines and institutions at the origin of wealth– is intimately related to what people know, and to the knowledge embodied in the available artefacts. The creation and allocation of economic wealth, therefore, co-evolves with the creation and diffusion of knowledge.

In the view of this thesis, understanding the dynamics, organization and viability of economies requires a better understanding of the organization of knowledge systems. The thesis proposes parsimonious theoretical and empirical models that help understanding the dynamics and organization of knowledge systems. In particular, knowledge networks are modelled as self-organizing two-mode networks. Two-mode networks are networks with two types of nodes, and links between nodes of different types. They are self-organizing in the sense that simple rules of evolution lead to a rich but patterned dynamics. For instance, as it is well-known in the literature on complex systems, positive feedbacks can generate power law (Pareto) distributions. In this thesis, a particular emphasis is put on non-linearities, typically logistic (S-shaped) curves, and how they affect self-organization and power law distributions. The main findings are: (i) stable power law distributions of ideas popularity naturally emerge from innovation and face-to-face diffusion; (ii) this dynamic is compatible with the observed (shifted) power law distribution of citations, and (iii) the generalized beta size-rank relation observed for patent classes can be explained by a slowdown in the growth of the number of classes. A general lesson from this work is that knowledge systems often exhibit non-equilibrium and non-linear dynamics, casting some doubts on their long term viability.

Chapter 2 reviews the literature on knowledge and development. Historically,

economics has always recognized the importance of knowledge in explaining individual and collective outcomes. In the second half of the XXth century, it was made explicit that there is no possibility of long-run per capita growth without productivity improvements. To construct an economic system capable of sustained growth, it is necessary for productivity growth to be an endogenous variable. Education, institutions and technological change have been identified as the main sources of productivity growth, and considerable effort has been devoted to understanding their origins. The endogenous growth literature has emphasized the mechanisms by which self-interested choices can be compatible with constant growth. But the importance of knowledge calls for a more explicitly micro- and meso-level analysis of knowledge systems. The distinctive feature of my approach resides in the use of theoretical and empirical models based on self-organizing networks. It is natural to use network theory for understanding knowledge systems. Knowledge is often tacit, and diffuses more easily face-to-face. In other words, knowledge flows take place on a network. As recent developments have shown, the structure of the network crucially influences the patterns of diffusion. In turn, innovation networks emerge from partner choice, which depends on the current allocation of knowledge. Moreover, besides agent-agent networks, the study of knowledge systems has made extensive use of network theory to analyze the relationships among knowledge artefacts themselves. Archetypical examples are scientific publications and patents. The references made by pieces of knowledge to other pieces form a citation network and reveal the overall structure of knowledge. More importantly, these references, by showing the relationship between new and existing knowledge, provide excellent information about the nature of growth in knowledge systems.

However, in spite of a large literature on social (agent-agent) and epistemic (idea-idea) networks, only little effort has been devoted to the formal study of their co-evolution. How does an agent-agent network relate to an idea-idea network? To answer this question, I argue that it is useful to formalize an "agent-idea" network. Two-mode networks, which have links only between nodes of different types, tie together the two one-mode (social and epistemic) networks. A contribution of this thesis (chapters 4 and 5) is to show that studying the dynamics of two-mode agent-idea systems allows us to relate, for instance, the structure of the social network to the structure of the citation network. Another contribution, chapter 6, uses two-mode networks to study the mesoscale organization of knowledge systems, that is, how ideas are classified into (an expanding set of) categories. To introduce the techniques used in chapters 4-6, chapter 3 presents in detail the most important models of growing one-mode and two-mode networks. In general, exact analytical results are hard, if ever possible to obtain. This is often the case in interaction-based

models, where one obtains macro-properties as emerging from self-organizing micro-behavior. However, different methods of approximation are available, and computer simulations allow us to check their validity, as well as to analyze more general cases inaccessible to pen-and-pencil methods. Chapter 3 puts a particular emphasis on a landmark model proposed by Herbert Simon in the early 50's to explain the emergence of power law (Pareto) distributions. This model can be seen as a self-organizing two-mode network, and is particularly well-suited to study agent-idea systems. Simon himself instantiated his model to author-paper systems, explaining the power law distribution of productivity in science (Lotka's law). However, it is clear and largely explained in chapter 2 that the creation of knowledge relies on previously acquired knowledge. There are no reasons to believe that in general, knowledge producers rely solely on knowledge that they have themselves produced in the past. Quite the opposite, the (mostly) non-rival and non-excludable nature of knowledge implies that it diffuses easily far away from its inventor. Simon's model does include innovation, but it lacks diffusion.

Chapter 4 proposes an original stochastic model in which agents generate new ideas and imitate existing ideas. There is a fixed number of agents, related by a social network. Agents learn random ideas of random friends, which creates a self-reinforcing dynamic in ideas' diffusion: the more well known is an idea, the more chances it has to be found in a random friend. This self-reinforcing dynamic would lead to exponential growth, but population is bounded, by assumption. Hence, the (expected) diffusion is logistic, in agreement with a large existing empirical and theoretical literature. If we wish to understand the diffusion of all ideas over time, we have to understand a stochastic discrete nonlinear dynamical system, which is also non-autonomous and with an increasing dimension. However, the *distribution* of ideas' popularity (the number of agents knowing an idea) can be characterized precisely, because it converges in the long run. The derivation of this result highlights the level at which self-organization takes place and gives rise to a "stable" state. Innovation and diffusion lead to a stable organization in terms of agents' knowledge overlap: the proportion of the ideas known by either of two agents which are known by both. This overlap among agents' ideas set determines and is determined by which ideas are learned. I present mean-field self-consistency equations that indicate the nature of this process and its convergence to a fixed point. At the steady-state, the distribution of ideas' popularity is a specific Generalized Hypergeometric Distribution, which tends to the well-known Yule-Simon distribution (a discrete power law) as the population size goes to infinity. Informally, it means that ideas' popularity is power law distributed, with a curvature of this power law due to population being bounded. Even though each idea is known by everybody

if we wait long enough, at every point in time there are a lot of ideas known by almost nobody and a few ideas known by almost everybody. I conclude the study by relaxing some assumptions. Changing the number of ideas that can be innovated or imitated per period does not change the results. However, changing the social network, which is assumed to be complete in the main developments of the chapter, can significantly affect the distribution of ideas' popularity. I show that a 2-regular one-dimensional circular lattice (a "circle" social network) leads to a geometric distribution. This suggests that distributions which have the exponential and the power law as limiting case (such as the shifted power law), constitute good expectations for the distribution of ideas' popularity, when that is unobserved but influences an observed quantity of interest.

Chapter 5 uses this result to explain the structure of citation networks. I extend the model of chapter 4 to explicitly model citations. It is simply assumed that, when new ideas (papers) arrive, they cite random ideas previously known by the inventor. Under this assumption, the exponent of the power law distribution of citations is the same as the exponent of the power law distribution of ideas' popularity, which is determined by the innovation-imitation trade-off. Hence this analysis departs from existing models, which explain the exponent of power law citation distributions by the relative prevalence of bibliographic and random search (reading papers cited by other papers, or found at random), These models are asocial, denying the collective nature of science, and do not account for the fact that attention is limited. The model of chapter 5 shows that limited attention and face-to-face diffusion can explain the observed patterns as well as backward bibliographic search models do. I also provide some evidence that, in spite of a pretty good fit obtained for the (shifted) power law predicted by both models (backward bibliographic search and social diffusion), the estimated parameters are not necessarily stable over time, indicating that the scientific citation networks studied (10 years of high energy physics) has a disequilibrium dynamics, due to the average size of reference lists being increasing over time. This disequilibrium dynamics may have important implications for the viability of knowledge systems, but the non-availability of long time series of the complete science system prevents a rigorous empirical analysis.

Chapter 6 analyzes the meso-scale dynamics of knowledge systems by studying the size of patent categories. How are patents distributed across technological categories? Answering this question requires an explicit consideration of the fact that new categories are created when genuinely novel technologies emerge. A large literature on category systems has found that the size distribution of categories is often very skewed, close to a power law. This observation for the size of biological genera even motivated U. Yule and then H. Simon to develop their models. When

size is power law distributed, the size-rank relation is also a power law, known as Zipf's law. In fact, the size-distribution of US patent *sub*classes is well fitted by a (shifted) power law. However, at the level of US patent classes, Zipf's law is clearly violated. Chapter 6 shows that the size-rank relationship is well fitted by a form close to a generalized beta distribution, which previous literature had identified in a variety of contexts, but which has not been explained by many theoretical models. To explain this pattern, I propose a new model which is a very simple non-linear modification of the Yule-Simon-Naranan principle (Naranan's model is the simplest version of the Yule-Simon mechanism). I assume that the number of categories grows as an asymmetric logistic (Richards) curve, while each category grows as in the Yule-Simon-Naranan framework. This result shows that when the growth of the number of categories tends to vanish as compared to the size of existing categories, a nonlinearity (in log-log terms) emerges in Zipf's law. The fact that this is the case in the US patent system may be a sign that radical innovations (new categories) tend to decrease as compared to incremental (same-category) innovations.

Besides the individual contributions of each chapter, a general message emerges from the thesis and is discussed in the conclusion: knowledge systems are self-organizing, but most likely they are in disequilibrium (the moments of the distribution of interest are changing over time). Moreover, significant nonlinearities have to be taken into account to accurately describe empirical patterns. Overall, this reinforces a complexity view of economics and knowledge systems, in which interactions, nonlinearities and disequilibrium dynamics are key components, but which can be usefully modelled by simple, self-organizing network models.

Chapter 2

Knowledge and development

Suppose that economic development, reflected in structural change, is driven by innovation and knowledge diffusion. Then theoretical economics should study the properties of innovation-imitation economies (as proposed in chapter 4), and explain empirical patterns found in various knowledge systems (such as citations among scientific publications or the size of technological domains, as proposed in chapters 5 and 6). This chapter explains that the knowledge-based nature of development is not an assumption, but a result of economic thought. The importance of technology, innovation and knowledge has always been at the heart of economics. Because economic wealth basically comes from people doing things, economic progress naturally comes from people doing things in a better way. People do things in a better way by using more or better knowledge, that they hold personally, collectively, or which is stored in institutions and artefacts. Of course, many factors interact to determine how individual productivity and collective organization evolve. But from a historical and dynamic point of view, an increasing division of labor in a world of expanding, improving varieties stands out as the main structural transformation incurred by economies along their path of development.

2.1   Historical background

It goes back to the classics to say that structural change and development are intimately related. Economic thinking, indeed, started off alongside the industrial revolution. Smith, Ricardo and Marx all made major contributions to the understanding of the fact that the movement towards specialization and diversification deeply impacts many aspects of human life and collective organization. From an analytical point of view, the concepts of specialization and diversification presuppose that there exists distinct types of activities. If that is the case, it is likely that at least some activities are dependent on some others, for instance for their

inputs. What exactly is the structure of these interdependencies? This question raised a lot of interest in the early and mid of the last century, alongside developments in national accounting and input-output tables. Hirschman (1958) insisted that industrial development should be thought with forward and backward linkages between sectors in mind. The Hawkins-Simon theorem (Hawkins & Simon 1949) gives condition under which an economy, represented by its input-output coefficients, can produce the right amount in each activity. General equilibrium theory (Walras 1874, Arrow & Debreu 1954) established conditions under which demand will be cleared on all markets simultaneously, and additional conditions under which welfare will be maximized.

Finding the conditions in which a complex system can or must be in equilibrium is informative if the system reaches its equilibrium (say, market clearing) much faster than this equilibrium is disturbed by external (or structural, but slowly moving) factors (Ando et al. 1963). This in essence legitimates neoclassical growth models, where an economy can grow, a deeply disequilibrating factor, while (almost) always being in equilibrium.

Growth models appeared early with Neumann (1945). Neoclassical growth models, such as the Solow-Swan and the Ramsey-Cass-Koopmans models, were developed in the 50's and 60's, embodying different amounts of general equilibrium theory, consumer rationality and capital accumulation dynamics. The development of econometrics increasingly allowed to test the correctness of certain assumptions, such as Cobb-Douglas production functions, and the validity of the predictions, for instance in consumer theory with Samuelson's (1938) revealed preferences approach. It is thus from the data that a bewildering fact arose. Solow (1957) estimated an aggregate production function and deduced that only 12.5% of the increase in labor productivity in the US (1909-1949) was due to an increased use of capital (a move along the production function). The 87.5% left (a shift of the production function), the "Solow residual", was interpreted as technical change. But in absence of a direct indicator for it, the Solow residual was dubbed "a measure of our ignorance" (Abramovitz 1956). Growth accounting was founded, and the second half of the XXth century largely confirmed the role of knowledge, embodied in capital, labor and institutions.

2.2   Knowledge-based growth and development

In the growth model of Solow (1956), there are decreasing returns to capital accumulation, leading to a steady-state level of GDP in the long-run, unless exogenous technological progress occurs. What can be added to a Cobb-Douglas production

function to obtain a steady-state rate of growth, that is, *endogenous growth*?

First, one can consider that most productivity improvements come from the skills of agents, obtained through education, on-the-job training, learning by doing and so forth. Lucas's (1988) model is centered around the accumulation of knowledge at the individual level, putting the focus on "human capital", and therefore on labor more than physical capital. In this case, the key decisions are education choices made by individuals. The emphasis on learning processes denotes a view of knowledge as difficult and long to acquire.

Second, one can instead see knowledge as perfectly codified, and easy to learn. Knowledge in this view can be thought of as a blueprint or capital vintage. Romer (1990) introduced a three-sector model in which R&D workers produce blueprints which are used by intermediate goods producers. Output is then determined by a production function in which there are an increasing number of blueprints. This increase in variety compensates decreasing returns within each type of intermediate goods.

Motivated by empirical questions on trade between asymmetric countries, Grossman & Helpman (1991) considered vertical instead of horizontal differentiation. They hypothesize that R&D allows products to improve on a quality ladder, introducing a creative destruction effect, when improved versions of products replace old ones. Schumpeterian growth theory (Aghion & Howitt 1992) built on industrial organization to study these business-stealing effects in detail.

An interesting debate within the endogenous growth literature is the presence of scale effects. Scale effects refer to a result present in Romer's model but not in the data. It says that the larger a country is, in terms of population, the higher should be its steady-state rate of growth. However, small countries do not in general grow more slowly than large countries (Jones 1995). Other models have been proposed which do not have a scale effect. Nevertheless, this shows that scale and aggregation are crucial issues.

In fact, knowledge is used by individuals, and most of their interactions take place at a local level. If knowledge is cumulative and localized, there should be self-reinforcing dynamics at the local level, suggesting that specialization patterns cannot be understood without reference to the idiosyncratic histories of particular places. Arthur (1994) and Krugman (1996) insisted that increasing returns generate self-reinforcing dynamics, and that under these conditions, self-organization leads to local places being differentiated in the knowledge space. An apparently innocuous "small historical event" can push a region into specific trajectories of specialization which are hard to escape. Important factors behind increasing returns are inter-sectoral (jacobian) and intra-sectoral (marshallian) externalities. These ex-

ternalities generate unique specialization patterns and clustering in space. It has
been suggested that for a large part, these externalities are localized due to the
stickiness of knowledge. In this case, it is not really the physical proximity that
generates externalities, but other types of proximity that favor knowledge diffusion
and creativity (Boschma & Frenken 2010). A case in point is given by patent cita-
tions. Jaffe et al. (1993) showed that patent citations are localized, with inventions
being more likely to cite other inventions born at the same place. Breschi & Lissoni
(2009) found that this is due to inventors mobility. Inventors change positions and
co-workers, contributing to knowledge diffusion, but most often mobility is local –
so are knowledge spillovers.

The literature on innovation systems (Freeman 1987, Lundvall 1992, Nelson
1993) has provided a theoretical framework to systematize the study of knowl-
edge flows and agents' interactions. The emphasis is on the interactive nature of
learning and on the complexity of the interactions among a diverse set of actors.
Private businesses, public agencies, universities, technological centers, and other
organizations engage in partnerships or exchange information and employees. The
institutions thus created can be formal or informal, and overall the organization of
the system that they constitute determines region (Cooke 2001), sector (Malerba
2002) or country-level innovation.

Innovative performance has been measured in various ways, such as total factor
productivity, learning curves, patent or scientific publication counts, and innovation
surveys. The multiplicity of methods reflects the difficulty to define innovation and
knowledge in an unambiguous and universal way.

2.3   The economics of knowledge

It seems hardly possible to define knowledge in a non-controversial way. How does
it differ from data or information? Does it exist independently of its carrier? Does
it really flow, or is it always recreated, reinterpreted, reintegrated into one's own
existing mental scheme? When and how can it be stocked and cumulated?

There are too many conceptual difficulties in the analysis of knowledge for a
short literature review to be fair. Pragmatically, and omitting from the history of
philosophy to the latest developments in artificial intelligence, this section recalls
some of the main properties of knowledge which make it a somehow peculiar object
for economic analysis.

Although the patent systems had been created in the XVIIIth century in Eu-
rope and the US, the economic literature on incentives to produce knowledge really
started out when Arrow (1962) and Nelson (1959) formulated it as a classical prob-

lem of externalities. Considering that knowledge is non rival and non excludable raises the problem of appropriation, and since then there has been many models aimed at understanding which institutions give the highest incentives to knowledge creation. For instance, patent race models study the effect of breadth, depth and length of private property protection on overall outcomes.

The assumption of non excludability of knowledge has been criticized on the grounds that it is not so easy to imitate knowledge held by others, even when it is codified in patents and scientific articles. First, it seems as if part of the knowledge people use to do things cannot be explained or codified. In Polanyi's (1962) well-known summary, "we know more than we can tell". A case in point that rigorous codification of procedures is not enough for effective transmission of knowledge was provided by Collins's (1974) study of the TEA laser: he found that no team could build one without the help of somebody having worked in a previously successful team. Whether or not everything *can* actually be codified and transmitted (is articulable), it is clear that not everything *is* codified. Thus agents make choices regarding codification and knowledge organization, storage, and access, and this choice is influenced by the technological and institutional context in which it takes place (Cowan et al. 2000). The point here is that we should expect a lot of heterogeneity, with different pieces of knowledge having different degrees of codification, and even identical pieces of knowledge being codified to different degrees depending on the context.

Furthermore, different individuals will have different abilities to acquire knowledge from a codified or uncodified source. This high individual heterogeneity can be explained by intrinsic abilities, but first and foremost, the ability to acquire a given piece of knowledge is determined by what an individual already knows. Knowledge is thus *cumulative*, and initial conditions matter. Often learning is cumulative because knowledge itself was created in a sequential way, with problems solved giving rise to new problems.

However, innovations and new ideas are sometimes disruptive. In some cases, new knowledge substitutes existing knowledge instead of complementing it. Schumpeter (1942) coined the term *creative destruction* to describe how innovations make existing products obsolete and can profoundly alter industrial organization. Knowledge cumulates until a point of rupture, when a breakthrough fundamentally changes which problems are interesting and which previous knowledge is useful to solve these new problems. These breakthrough are rare and their impact lasts over time. Hence they create strong non linearities in economic change. Radical innovations spark a wealth of other innovations early in the cycle, creating innovation clustering in time. The non linear evolution of knowledge and technologies suggests long waves

à la Kondratiev (Silverberg 2002).

The cumulativeness of knowledge is often associated to the idea that knowledge is created by recombining existing knowledge. This combinatorial view has an optimist corollary, as the number of possible combinations between ideas rises extremely fast with the number of ideas, making long run growth possible (Weitzman 1998). But the cumulativeness of knowledge also lead to less optimistic conclusions. If knowledge really grows, then there is more and more to learn. If new knowledge does not make previous knowledge easier to learn, newcomers should devote more efforts to reach the frontier. In Jones's (2009) telling metaphor, it becomes harder to climb the giant's back as the giant grows[1]. As evidence for this "burden of knowledge", Jones (2009) shows that the age at greatest invention increased over time, as did time to PhD graduation. For him, this increasing burden of knowledge is responsible for an increasing specialization and propensity to work in teams. The interesting point is that the organization of knowledge systems (the structure of who knows what) cannot be stable as knowledge grows.[2]

To conclude, knowledge is very heterogenous, and it is difficult to make general claims. Knowledge is often cumulative but sometimes a breakthrough displaces old knowledge. Knowledge is often non-rival but sometimes too hard to imitate. Knowledge accumulation creates possibilities for further progress but puts an increasing burden on newcomers. Moreover, knowledge is learned by individuals but it is often stored in organizational routines or institutions. One can then approach knowledge as rules which are created, diffuse and are displaced by new rules in a perpetual cycle of variation-selection-retention.

2.4   Evolutionary economics

Evolutionary economics is born from the idea that economic forces can be understood by analogy with evolutionary biology. This is an old idea, and economics and biology have repeatedly exchanged metaphors and mathematical tools (Hodgson 2006).

In a light version of the argument,[3] the pressure induced by competition among

---

[1]The original quote from Isaac Newton in 1676 is "If I have seen further it is by standing on the shoulders of giants."

[2]The relation between knowledge growth and the organization of knowledge systems is the main point of inquiry of this thesis. In Chapter 4, I propose a model which exhibits self-organization into a stable state at the level of the distribution of ideas' diffusion. But in chapter 5, it will be noted that citation distributions have a shape (and mean) changing over time. In chapter 6, I have to introduce a strong non linearity in growth to explain the size-rank relationship of patent classes.

[3]There are a number of controversies about the link between biology and economics. The map of the human genome and the availability of large datasets drive current work in evolutionary

firms in a market is analogous to that faced by biological organisms competing to survive and reproduce in a world of limited resources. Market forces act as a selection device, and only the most profit-maximizing firms can survive, justifying the profit-maximizing assumption (Alchian 1950).

At a closer microeconomic inspection, however, evolutionary forces are better thought at the level of behavioral routines (Nelson & Winter 1982). When a new type of behavior appears (*variation*), if it is successful (it "increases fitness" in the sense of providing additional utility), it will be *selected*, and its imitation in the population can be thought of as *retention*. In an evolutionary world, there are many rules constantly being created, and eliminated or selected. What is needed for an evolutionary theory of economic change is a theory of behavior and rationality based on rules. This theory was provided by Herbert Simon, under the rather misleading label of "bounded" rationality.

Simon (1947), equipped with microeconomics principles, studied decision making of city planners. Deceptively, they did not seem to be checking all alternatives, formalizing their preferences, and finding an optimum. Instead, they followed rules of thumb that they would adapt to changing circumstances. The birth of computing around this time greatly helped Simon formalizing his ideas on heuristic problem solving (Newell et al. 1972), and many models embodying some degrees of bounded rationality have been introduced (Simon 1956, Dosi et al. 2006, Kirman 1993, Young 2001). Taking the idea of a *satisficing rule* seriously, Nelson & Winter (1982) proposed an evolutionary metaphor of how routines are discovered, imitated and selected. They showed how procedural rationality explains industrial dynamics, pioneering agent-based modelling on classical questions of industrial organization, such as the number of firms, technology races, and demand dynamics (Nelson & Winter 1982, Malerba et al. 1999). Dopfer et al.'s (2004) discussion of economic evolution is a plea for a mesoscale understanding of economic dynamics. Microeconomics should explain us how rules are generated and imitated. Macroeconomics provides a global understanding of the relationship between aggregated variables. The interesting dynamics is at the mesoscale, which should keep track of innovation and diffusion at a semi-aggregated level, by considering relevant averages and distributions. In this framework, rules exist somehow independently of their users/carriers. But in practice, what matters economically is knowledge that is effectively used.

The relationship between knowledge and productive labor is a complex one. Agents learn by doing, and competition induces them to do preferably what they are good at, which is what they know the most about. This two-way relation between

---

psychology and sociobiology, which tries to settle the debate between the influence of genetic and cultural factors on behavior.

the division of labor and the division of knowledge is dramatically affected by the disruptive forces of innovation, and raises important issues of coordination (Becker et al. 2007, Marengo & Dosi 2005). The focus on actual use of knowledge also suggests bottom-up approaches to knowledge-based development. Traditionally, for instance according to Vannevar Bush's doctrine for Science and Technology Policy, knowledge generation and use is seen as a top-down process, with knowledge being created by science and flowing down to its productive users. This linear model of innovation has been largely criticized and amended with virtually every possible feedback loop (Kline & Rosenberg 1986).

The discussions on knowledge tacitness and bottom up innovations is interesting in the context of developing countries. An older literature on appropriate technologies, which was criticizing top-down "technology transfer", is now rejuvenated with "inclusive innovations" or innovations for the "bottom of the pyramid", highlighting that knowledge-based development is not only (if at all) a process of imitation and catch-up but the construction of local capabilities and institutions conducive of creative and cooperative interactions.

To understand this process on the basis of theoretical models, it is necessary to start with a simplified view of what is knowledge. Since this choice is delicate and influences the results and focus of study, the next section discusses some of the possible choices in that matter.

2.5   Modelling innovation and imitation

The literature on innovation and imitation is huge (Hall & Rosenberg 2010, Rogers 2010). Here I describe different approaches to model knowledge growth, including knowledge as a quantity, knowledge as ideas (possibly with an underlying topological structure), knowledge as problem solving, and knowledge as learning about a state of the world.

A common approach to modelling knowledge is to assume that it is a level of expertise. In this case some agents have more knowledge than others, and diffusion is a catching-up process. Innovation is modelled independently, with some agents pushing the frontier (Ehrhardt et al. 2006). When there exist different types of knowledge, one can study barter exchange. In this case, it can happen that knowledge does not completely diffuse, because in some parts of the network, agents do not have an interest to exchange (Cowan & Jonard 2004). This depends on the exact structure of the network on which diffusion takes place.

A different modelling approach is to consider discrete ideas. In this case, knowledge can still be seen as a quantity, simply because the number of ideas can be

counted. But now it can be that two agents with the same amount of knowledge have something to exchange, because they have the same number of ideas, but not the same ideas (as when one considers that there are different *types* of knowledge). Discrete "ideas" have sometimes been thought of as blueprints (Romer 1990) or production recipes (Jones 2005).

A more elaborated approach is to assume that knowledge itself has an underlying structure, and this structure is intimately linked to time, due to the cumulative, autocatalytic nature of knowledge (Vega-Redondo 1994, Carayol & Dalle 2007). Although making assumptions on the underlying structure of knowledge may seem delicate, the availability of large datasets on citation and co-classification networks makes it possible to actually observe how different branches or fields of knowledge are related and co-evolve.

Saying that different pieces of knowledge are related is only one step from saying that knowledge is a complex system of interacting elements, and that if it is, we might expect a range of nonlinear phenomena to take place. This was illustrated by Silverberg & Verspagen (2005) who proposed a percolation model for innovation. Innovators explore a technological space which is conceived as a topological structure given by Nature (here, a lattice).

Another approach to modelling knowledge starts from the idea that knowledge creation is essentially a process of problem solving. Technological and scientific evolution can be described as new solutions giving rise to new problems (Arthur 2009). The problem solving perspective allows us to acknowledge that in general there are competing problems to be solved, and competing methods to do so. By analogy to Kuhn's (1962) "normal science", Dosi (1982) identifies technological paradigms with what in a particular context is judged as "relevant" problems to solve and ways to solve it.

A crucial step in the modelling of problem solving was made possible by Kauffman's (1993) work on evolutionary biology. Kauffman (1993) considered the fitness attained by a species which incurs random mutations. Assume that a species is defined by a vector of elements, each element can be 0 or 1. Each element has epistatic relations to only K others,[4] that is, the fitness of an element is the same whatever the state of an unrelated element is. On the other hand, the fitness of an element being on say, 1, depends on a related element being on 0 or 1. Fitness values are otherwise random, and the overall fitness of the organism is the average of the fitness of its elements.

This way of computing fitness implies that very different system states (vectors) can have high fitness, but the states in-between (which random mutations should

---

[4]This can be thought as a K-regular directed graph. When $K = 0$ or $K = N - 1$, it is unique. Otherwise one of all the possible K-regular directed graph is chosen.

follow to go from one to the other) may have lower fitness. Hence evolution might not be able to explore the whole state space. In, fact, evolution gets stuck whenever it is in a local optimum. At this point, any single mutation would lower fitness, and since multiple mutations at once are not allowed, this is an equilibrium.

The NK model described above provides a powerful representation of a hard problem that cannot be solved optimally by heuristic search, providing a good framework to study how finding good solutions depends on problem complexity and the search algorithm. In this interpretation, an agent has to design a technological system by choosing a value for each element (Frenken et al. 1999). Of particular interest in the technology literature has been the emergence of modular technological structure (Baldwin & Clark 2000), a phenomena that has wide ranging consequences on the organization of innovation and on market structure. Using a NK model, Frenken et al. (1999) showed that the depth of search necessary to find the global optimum decreases with the decomposability of the problem. A problem is decomposable when the fitness of some groups of elements is independent of outside-group elements, and can thus be optimized independently. Nearly-decomposable systems, which have a few intergroup relations, can also be searched effectively and have evolutionary advantages (Frenken et al. 1999, Simon 2002, Marengo & Dosi 2005). In another use of the NK framework, Page (2008) studied how the diversity of teams influences problem solving performance.

Finally, knowledge growth has also been modeled as a learning process. Generally, one assumes that there exists a (possibly changing) true state of the world, and agents form and revise beliefs about it. One is then interested in the convergence of individuals' beliefs to the truth. How beliefs are formed and revised depends on the particular model being studied. Learning models are thus extremely numerous and diversified, also due to the interest that they have received from different communities for different purposes. Theoreticians of games and general equilibrium use learning models to find the conditions under which equilibrium or consensus will be reached (Marimon 1997). Evolutionary economists use learning models that convey a large number of bounded rationality principles (Dosi et al. 2006). Computer scientists use learning models in artificial intelligence (machine learning). In biology, learning in evolutionary games has received a lot of attention recently. More generally, learning models for populations of agents have been well studied, and tend to be particularly appropriate to study the effect of an underlying interaction structure (Bala & Goyal 1998, Golub & Jackson 2010).

Of course, the link between an increasing number of "ideas" and economic development is complex and by no means automatic. In fact, some ideas turn out to make people rather worse off, or to benefit a few to the detriment of others. At the

very least, we should expect that even "positive" technologies create inequalities, as is clear from the wide literature on diffusion. By definition, innovation implies a leader-follower structure, in the sense that the innovator (or the team) is always the first to know, and is followed by others, who catch-up. However, when there are different types of knowledge, or when ideas are modelled as discrete objects, agents can be both leaders and followers.

This thesis takes the view that approaching this ecology of creation and diffusion formally is best done using the theory of networks, by putting interactions at the center of the analysis. It is part of a general research program on agent-based models and complexity in social sciences which has been driven by the advent of computers. In this context, it was discovered that simple rules can generate complex and interesting patterns of behavior. Complex behavior implies that it is very hard and sometimes impossible to predict mathematically how such systems behave. But simple rules imply that these systems are easy to simulate, and for this reason simulations have been heavily used to understand even toy models (Schelling 1960, Epstein & Axtell 1996, Axelrod 1997).

Agent-based models have also been studied using mean-field theory and a range of other techniques of approximation. The computer keeps track of what happens to every agent; and the modeler constructs an approximate mathematical model of the computer model. The mathematical predictions can then be checked against the simulations. This practice allows us to go deep into the microdynamics thanks to the computer while having a mathematical understanding of the meso- and macro-dynamics. This approach will be repeatedly used in this thesis.

Perhaps even more important than agents are the *relations* among the agents. It has been found throughout the last 60 years or so that graphs are very appropriate mathematical objects to study social (or others) interactions empirically and theoretically.

A graph G is a set of vertices or nodes, and a set of links or edges between pairs of nodes. It is usefully represented by an adjacency matrix A in which the entry A_{ij} shows the relationship between i and j. Generally, A_{ij} takes the value 0 or 1 to express whether or not "i is friend with j". It can also be an arbitrary scalar, indicating the strength of the relationship (edge weights). If edges are undirected, A is symmetric. Graphs can be represented by drawing curves between points. One has to choose first where to draw the points (the nodes), and then edges can be added. The graphical representation of complex systems as graphs is intuitive and widely used qualitatively. But in spite of great progress in visualization techniques, one often resort on more concrete mathematical analysis to extract meaning from networks.

2.6   Social network analysis and complex networks

There does not really exist a point in time at which networks were invented or discovered – for obvious reasons social and kinship networks are older than homosapiens. But there certainly is a beginning to the technical creation of nets (e.g. for fishing) and a long sequence of trial and errors in organizing paths, for pedestrians, water and other infrastructures. As technology, science and society grow in complexity, networks are more and more apparent and their organization more and more consequential (Castells 2011).

From a mathematical point of view, the history of graph theory starts in Königsberg, where Euler concerned himself with the puzzle of travelling the whole city by walking each of its seven bridges once and only once. The nature of the problem lead him to formalize pieces of land as dots and bridges as lines between the dots. Graphs were born and they quickly delivered their first result. To travel Königsberg as required by the puzzle, the nodes along the path taken (excluding the first and last) must have an even degree (an even number of edges attached to them), so that one can enter and then leave. Because more than two nodes had an odd degree in 1736's Königsberg, there was no path as required by the puzzle.

Besides the work on discrete structures by the mathematical community (Berge 1989, Bollobás 1998) sociologists have been at the forefront in the use of graphs. The idea of an underlying "social structure" is as old as sociology, but Social Network Analysis, as it came to be known, started out in the 50's. By the mid-90's, the field had produced a wide range of methods (Wasserman & Faust 1993) soon followed by computer programs, leading to a large diffusion of the techniques.

Among many insights from this large literature, the tension between triangular closure and structural holes is worth noting. Triangular closure (or clustering, or cliquishness) is the propensity of two friends of ego to be friends with each other. Triangle closure can be interpreted as a sign of high social capital, in the sense that it generally indicates that agents form tightly connected subgroups. On the other hand, when friends of ego are also friends of each other, information tends to be redundant. Friends from another group, who are not connected to other friends of ego, and thus supposedly only "weak ties", tend to provide different information: hence the "strength of weak ties" (Granovetter 1973). In addition, observe that when two friends of ego are not connected, ego is in between, therefore it has some power over information flows in the network (Burt 1995) and can capture more rent.

The example above only hints to the rich theories that can be embodied into formal propositions about network structure. Another legacy of XXth century social network analysis is a set of centrality measures, each capturing a different interpretation of what it means to be prominent, powerful or "central" in a network.

For example, the degree only counts the number of friends. Closeness evaluates the average geodesic distance (shortest path) with all others. Betweenness counts how many pairs of players have a shortest path through ego. Bonacich's (1987) centrality takes a linear algebraic approach. Assuming that the centrality of ego is the sum of the centralities of her friends leads to an eigenvalue equation. Since the adjacency matrix is non-negative, the Perron-Frobenius theorem ensures that there exists a positive solution vector for the highest positive eigenvalue.

Besides the progress driven by empirical work, mathematicians have been interested in random graphs. In particular, Erdős & Rényi (1960) studied a model in which each possible edge exists with independent probability p. They showed that the size of the largest connected component is a very special function of p. Below a critical p_c, the largest connected component has very high probability of being very small, and above the critical probability it has a very high probability of being very large. The system exhibits a phase transition at the critical point at which nodes are connected to 1 other node on average.

Random graphs are also used as null models in empirical work. Often one goes further than simple Erdős & Rényi (1960) graphs and also imposes a given degree sequence (the degree of each node) that the random graph must have. To create a random graph with a given degree sequence, one needs to count up all the possible graphs satisfying this constraint, and then choose one of these graphs at random. In practice, it is computationally too expensive, and stub-matching algorithms are used to create a graph which respects the constraint. The analysis of random graphs revealed that the sheer number of total possible graphs makes it difficult to count and classify graphs, with some graphs sharing a common property but being completely different with respect to another property. When some of these properties are desirable or, on the contrary, need to be avoided, it is very difficult to find an optimal graph.

Watts & Strogatz (1998) made a methodological and substantive contribution to this issue by proposing to study two properties in a family of random graphs interpolating from a lattice to an Erdős-Rényi graph. In a lattice, such as a graph where agents are organized in a circle and are connected to a few others on each side, the distance between agents is pretty high because one needs to travel the whole lattice to go from one side to the other. On the other hand, clustering[5] is high in a

---

[5]Watts & Strogatz (1998) define clustering at the local level, that is, for each agent, how many pairs of friends of ego are also friends of each other. This individual clustering can then be averaged, which is meaningful in their context. However, it must be noted that individual clustering is ill-defined for agents with no or one friend. Moreover, if we want the *probability* that two random friends of a random node are friends, we need to compute clustering at a global level, by counting the number of connected triples (three connected nodes) and checking what is the share of those that are closed (i.e. "triangles", which are full subgraphs of three nodes).

lattice such as the one described above, where agents are densely connected locally. How can it be that social networks seem to be made of tightly connected groups, and at the same time empirical claims are made about the "small-world" phenomena, according to which every two persons on earth are connected by a chain of at most 6 acquaintances?[6] Watts & Strogatz (1998) showed that by rewiring only a few edges of the lattice, clustering would not drop as fast as path length. That is, starting from the lattice, if we pick an edge, detach it on one side and put it back to a node chosen at random, clustering has decreased locally (for the few nodes involved where the link was detached from), but path length has decreased for potentially many nodes who now find it easier to reach the other side of the lattice by going through the rewired edge, which plays the role of a shortcut. By rewiring every link the graph would become completely random. In this case path length would be low, but clustering would be low as well.[7] Hence it is between a completely ordered (lattice) and a completely random network that we can find a system with both low path length and high clustering. This result received a lot of attention after the literature on the edge of chaos and self-organized criticality.

Another example of the use of random graphs is the modularity maximization criterium for community detection. Community detection in networks seeks to divide the nodes (or edges in some cases) into possibly overlapping subsets of nodes such that, loosely speaking, most interactions are within communities, and there are only few inter-communities interactions. To see how crucial it is to uncover the community structure of networks, consider the Simon-Ando theorem (Simon & Ando 1961). They consider nearly-decomposable dynamical systems, that is, with variables that can be organized in almost independent groups. They showed that convergence can be decomposed into short-run convergence of individual subgroups of variables (dominated by the largest eigenvalue in this subgroup), and long run convergence of the blocks to the unique steady state (dominated by the largest eigenvalue of the whole matrix). Hence a dynamical process taking place on a network is greatly affected by the degree of modularity, or near-decomposability of the network.

Often, modular systems are also hierarchical. This implies that the network can be meaningfully split into subnetworks, and in turn these subnetworks can be decomposed. In some nice cases where decomposition is self-similar, one can see the fractal structure of multi-level systems. More generally, the emergence

---

[6]This figure is of course debated. See Milgram (1967), de Sola Pool & Kochen (1979) and the review of Schnettler (2009).

[7]In practice Watts and Strogatz normalized their measure of clustering and path length with that of Erdős-Rényi graphs. Since edges exist independently in the Erdős-Rényi model, the probability that two friends are friends of each other is simply the probability that two nodes are friends of each other, which is fixed by definition of the model.

of the complex networks literature has allowed great progress in the modelling of multilevel systems. On the one hand, the study of community structure has been the source of numerous ideas about the hierarchic nature of complex systems. In fact, to find an optimal way of splitting a set of nodes into homogenous groups, algorithms often construct a whole hierarchy of subsets (a dendrogram), and use a given criterium such as modularity to choose which layer of the dendrogram is the best. The dendrogram delivers information about the hierarchical organization of the network, and some authors have turned the community detection problem into one of revealing the multilevel organization of networks (Granell et al. 2011). On the other hand, another branch of the complex networks literature has chosen to extend one-mode networks to more complicated structures, so as to deal with multiple interacting networks (Kivelä et al. 2013). Two-mode networks are perhaps the simplest of these extensions, and potentially allow to integrate easily the two one-mode networks on each side of the two-mode network, as shown concretely in chapters 4 and 5.

Another connection between two-mode networks and multi-level systems can be seen using hypergraphs. A hypergraph is a set of nodes and a set of hyperedges. Each hyperedge is a subset of nodes. This can be fully represented by an incidence matrix B where $B_{ij} = 1$ if the node i is contained in the hyperedge j. Hypergraphs are now sometimes used to study multimode networks. In this case, the nodes of an hyperedge can (and in most applications must) come from a different mode. For instance, in a three-mode network of tags, users and photos, Zlatić et al. (2009) define a hyperedge as a user tagging a photo. From this example, we sense that the emphasis is on agent-artefact interactions. In many cases, these agent-artefact interactions can be thought of as events occurring over time. There is a literature on hypergraphs (Seidman 1981) and simplicial complexes (Atkin 1977) which identifies precisely which information is lost when we require all relations to be among pairs of agents only, instead of more general group-level relations. The theory of self-organizing multi-mode networks allows to incorporate old ideas about multi-level systems into the framework of network science.

2.7 The economics of networks

The economics of networks is now well established, with a number of monographs and textbooks (Goyal 2007, Vega-Redondo 2007, Jackson 2008, Easley & Kleinberg 2010). Besides the connection between input-output and networks, economists have traditionally been interested in networks in the context of industrial organization. Many industries that require a large infrastructure to operate are characterized by

economies of scale leading to natural monopolies. In these cases state intervention is required, either as direct owner or through price regulations (such as the Ramsey-Boiteux rule). Moreover, there exist many industries characterized by demand-side externalities, known as network effects. This situation arises, for instance, when there are competing standards. Standards create an incentive for newcomers to adopt the standard which is currently the most popular, leading to monopoly. Hence, if by historical accident the worse standard is more widely adopted early on, the best standard will not be adopted even in the long run, a situation of lock-in (David 1985, Arthur 1989). This analysis lead to a body of work on standards and compatibility issues (David & Greenstein 1990), technology choice (Cowan 1991), and industrial organization (Economides 1996, Shy 2001).

In spite of some attempts to study the topology of the underlying network (Economides 1996), this early literature did not systematize the use of graph theory. Game theorists instead started to use graphs along different directions. Models of network formation are often divided between access (one-sided decision) and connection (two-sided) models. For instance, Jackson & Wolinsky (1996) studied a connection model in which externalities decay along geodesic distance, and prove that (pairwise) stable networks are not necessarily efficient. Bala & Goyal (2000) studied an access model and again found a possible discrepancy between stable and efficient networks. Other models have taken the structure of the network as given, and study the effect of network structure on collective outcome, typically using a coordination game (Morris 2000). Vega-Redondo (2007) provides a framework with endogenous strategy and partner choice, characterizing the coevolution of networks and institutions. A specific type of network formation particularly relevant for this thesis is based on growth. Instead of considering all players and ask who wants to connect to whom, one starts with few players, and ask to whom the newcomers want to connect to. These growing network models are central to the thesis and will be reviewed in the next chapter.

2.8   Power laws and stochastic growth models

In this thesis, networks and power laws are central topics. This is due to the nature of knowledge systems. If we denote by $p(k)$ the fraction of observations with value k, a power law is a relation $p(k) = Ck^{-\gamma}$ where C is a constant that must ensure correct normalization of the probability distribution[8]. Typically power law behavior ("scaling") is observed only within a certain range. Power law (or

---

[8]This constant differs depending on the data being discrete or continuous, and depending on the admissible range of k.

Pareto) distributions are found for many quantities of interest, in particular size-distributions. For instance, the empirical observations of chapters 5 and 6, citation distributions and patent category sizes, are suggestive of power laws or "similar" distributions. Under one form or another, this distribution has been observed in, *inter alia*, the number of times single words are being used in a text (studied by Estoup), the size distribution of cities (studied by Zipf), the distribution of income (studied by Pareto), the distribution of the number of papers published per author (Lotka's law), and the size distribution of genera (observed by Willis), which was explained by a mathematical model in Yule (1925). This universality is interesting and has stimulated theoretical models to explain these power laws. The model of chapter 4 is an extension of a core model in this area due to Simon (1955). This model gained renewed interest after its rediscovery (driven by empirical observation) in the context of networks by Barabási & Albert (1999). Simon's model is different from Barabási & Albert's model, but both are growing networks.

This thesis aims at understanding the self-organization of knowledge systems, but the models of two-mode network growth presented here are relatively general, and relate to several other branches of the economic literature which has used them to understand wealth distribution, firm size distribution, and, more recently, product diversification or other quantities in international trade. Pareto discovered his eponym law using income data, and many stochastic models for the emergence of power laws were designed with income distribution as their main application (e.g. Champernowne (1953), Shorrocks (1975)). Regarding firm size distributions, the major early work is of course Gibrat (1931) who showed that multiplicative shocks generate a lognormal distribution: if firms have an expected growth rate independent of their size, the central limit theorem can be invoked to show that the log of the size is a Gaussian distribution. Simon (1955) obtained a power law, in fact a Yule-Simon distribution, by adding an entry process. Steindl (1965) and Ijiri & Simon (1977) contributed other, more elaborated models but the lack of very large datasets prevented rigorous empirical testing. Axtell (2001) studied a large dataset and found that the power law gives a better fit than the lognormal in the case of city sizes. Gabaix (1999) modified Gibrat's principle by assuming a random walk with a lower reflecting barrier, and find that it leads to a Pareto distribution. Fu et al. (2005) proposed a model which reproduced the tent-shaped distribution of growth rates. Bottazzi & Secchi (2006) used a stochastic growth model to study diversification. More recently, stochastic growth models for the emergence of power law distributions have been included in neoclassical economic models, following the upsurge of interest on the causes and consequences of firm heterogenity (Luttmer 2007). Stochastic growth models have also been applied to microeconomic data in

international trade, such as the two-mode networks of which firm or which country exports which products (initiated by Hidalgo et al. (2007)). For instance, Armenter & Koren (2014) proposed null models, while Riccaboni & Schiavo (2014) and Chaney (2011) introduce more microfounded principles. Finally, there has also been applications in economic geography, where they are particularly well suited to model the positive feedbacks typical of spatial agglomeration processes (Arthur et al. 1987). For instance, Frenken & Boschma (2007) derived simultaneously the distribution of city and firm sizes, and Bottazzi et al. (2007) proposed a testable model of spatial agglomeration under positive feedback and entry-exit of firms. More generally, stochastic birth and death models with an explicit account of the topology of the underlying spaces (geographical, social, cognitive...) seem particularly appropriate for constituting the field of "organizational genealogy", with the aim of encompassing industrial dynamics and economic geography in an evolutionary framework.

2.9   Conclusion

This chapter aimed at backing-up the main premise of this thesis: economics is fundamentally about the creation and diffusion of knowledge. Knowledge has to be put to use to create economic value, but it is a pre-requisite and determines most of the economic system dynamics.

Neoclassical knowledge-based growth models provide an encompassing framework relating knowledge growth with welfare, but at the cost of delicate assumptions, such as substantive rather than procedural rationality. Evolutionary economics has proposed numerous aggregated and agent-based models with rich knowledge - based dynamics, but a synthesis is still missing. Knowledge itself is hard to comprehend as a homogenous category, which is challenging for the design of theoretical models. In fact, an issue in the literature on interaction-based models of knowledge dynamics is that it is difficult to model innovation and diffusion in a meaningful yet parsimonious way. Hence, one often has to resort on simulations to understand the model, making comparison with data much harder. This thesis proposes agent- or interaction-based models and devotes special effort to finding analytical approximations, leading to straightforward empirical estimations.

In chapter 4, I propose a new framework to study analytically the interplay of innovation and imitation in interaction-based models. I show that the evolution of a system in which ideas are created and diffuse can be formulated as a self-organizing two-mode network. In line with a large empirical and theoretical literature, the expected diffusion of each idea follows a logistic (S-shaped) curve. This dynamical

system is a non-linear extension of a well-known model (Simon 1955), and can be relatively well characterized in spite of its dimension (the number of ideas) being increasing. The distribution of ideas' popularity is an extension of the Yule-Simon distribution, and can be thought of as a power law with a significant upward or downward curvature. The innovation-imitation trade-off determines the slope of the power law.

In chapter 5, I propose to extend and modify the theoretical model of chapter 4 to bring it to the data. Science is a knowledge system in which new ideas are created and exchanged. I consider the distribution of citations received by scientific publications, which is often well fitted by a shifted power law. The literature explains the value of the power law exponent by the relative rates of "random" and bibliographic search (whether cited papers are found at random, or by reading the bibliography or other papers). Based on theoretical knowledge obtained in chapter 4. I propose a different interpretation: the shape of the citation distribution mainly reflects the way in which attention is traded-off between innovation and imitation. It is also influenced by the structure of the social network on which diffusion takes place.

While chapters 4 and 5 start from theory to go to the data, chapter 6 takes the opposite route. There, I find that the size-distribution of technological subcategories (US patent subclasses) follows a relatively expected distribution (not far from Zipf's law). But at a higher level (classes), the empirical picture cannot be explained by existing models. This motivates the introduction of a new model for this class of phenomena. To be compatible with the empirical evidence, the new model posits that the number of categories follows an S-shaped curve.

Chapter 3

Models of growing networks

In this chapter, I review simple models of growing (or self-organizing) networks. There are two important distinctions to be made. First, there are one-mode and two-mode networks. Two-mode networks have two types of nodes, for instance authors and papers, with a link between an author and a paper meaning that this particular author wrote that particular paper. Second, there are two major techniques of analysis, mean-field continuous approximations and master equations. As far as Simon's (a two-mode network) and Barabási & Albert (BA)'s (a one-mode network) models are concerned, both techniques work on each model. However, later in the thesis, master equations become untractable and it is thus necessary to present both approaches. Consistent with original works, the BA model is presented using the mean-field continuous method, and Simon's model using the master equation.

This chapter also contains new results and original derivations of known facts. Section 3.1 reviews growing one-mode network models, and shows precisely the link between average degree and the parameters of the shifted power law in shifted attachment kernel models. Section 3.2, which reviews Simon's (1955) model, includes an original derivation of the Yule-Simon distribution and proof of proper normalization, to introduce some techniques used in chapter 4. While section 3.3 only reviews other models of self-organizing two-mode networks, section 3.4 proposes an encompassing model of two-mode network growth by addition of (complete) subgraphs. It generalizes Simon's model and its extension by Ramasco et al. (2004) for collaboration networks and, as in section 3.1, it clarifies the respective roles of initial and acquired degrees, the latter being preferentially attributed and the former not.

3.1   Models of growing one-mode networks

Barabási & Albert (1999) proposed a model to explain their observations of the

structure of the World Wide Web ("the BA model"). The nodes are web pages and the links are hyperlinks. They found that the degree distribution is a power law, that is, the share $p(k)$ of pages with k links is proportional to $k^{-\gamma}$. They developed a model to explain that this power law distribution is due to two main factors: *growth* and *preferential attachment* of new pages to popular existing pages.

The model is a growing undirected graph. At each period, a new node arrives in the network, and connects to m other nodes. The m chosen nodes are selected with probability proportional to their degree, hence the expression preferential attachment.

To find the degree distribution, Barabási & Albert (1999) proposed to study the evolution of individual node's degree using a deterministic, continuous time, continuous degree equation. In this approximation, the degree of node i, denoted k_i, evolves as

$$\frac{dk_i}{dt} = m\frac{k_i}{\sum_i k_i} = \frac{k_i}{2t}, \tag{3.1}$$

where the last relationship follows from the fact that since m edges are added at each period, there are $2mt$ degrees (i.e. end of edges) in the system. The solution of (3.1), with the initial condition that i is born at time t_i with degree m is

$$k_i(t) = m\left(\frac{t}{t_i}\right)^{1/2}. \tag{3.2}$$

Using this information for all i one can find the share of nodes with a certain degree. Using the cumulative distribution function, we have

$$p(k_i < k) = p\left(m\left(\frac{t}{t_i}\right)^{1/2} < k\right) = 1 - p\left(t_i < t\left(\frac{m}{k}\right)^2\right). \tag{3.3}$$

Since there is exactly one node born at each period $1 \ldots t$, the t_i's are uniformly distributed, that is, $p(t_i < Y) = Y/t$. Using this, (3.3) becomes

$$p(k_i < k) = 1 - \left(\frac{m}{k}\right)^2. \tag{3.4}$$

Note that (3.4) simply shows that, under these approximations, the fraction of nodes with degree less than k at time t is the fraction of nodes born after the node who has exactly degree k at time t (Jackson 2008). From (3.4),

$$p(k_i = k) = \frac{dp(k_i < k)}{dk} = 2m^2 k^{-3}. \tag{3.5}$$

Equation 3.5 is a power law with exponent 3, so the BA model predicts a universal

exponent. This is both a good and bad feature of this model. This is good because empirical data shows that the exponent is often not too far from 3, and this is explained in the simplest way by the BA model. However, the departure from 3 is in some cases quite significant, which triggered additional research modifying the BA model to obtain a tunable power law exponent. Furthermore, the derivations above rely on continuous time and continuous degree approximations, and keep track only of the expected degree. To obtain a better characterization of the BA model, one can use master equations, a task carried on simultaneously by Dorogovtsev et al. (2000) and Krapivsky et al. (2000). However, not all extensions of the BA model can be solved using master equations. For this reason, I first show the different extensions using the mean-field continuous approximation of BA, and I will introduce master equations to discuss two-mode networks in the next two sections.

Dorogovtsev et al. (2000) start from the insight that in the BA model, edges should be seen as directed, running from the newborn node to the existing ones. Suppose that we want to know the distribution of indegrees. According to the BA, model nodes receive edges with probability proportional to their total degree, which is their indegree, that changes over time, plus their outdegree, which is m throughout their life. Denoting by q_i the indegree of node i, the BA model implies

$$\frac{dq_i}{dt} = m\frac{q_i + m}{\sum_j [q_j + m]}.$$

Dorogovtsev et al. (2000) propose to generalize this by assuming[1]

$$\frac{dq_i}{dt} = m\frac{q_i + a}{\sum_j [q_j + a]},\tag{3.6}$$

where a is the *initial attractiveness* of a node. If this term was 0, new nodes would not be able to receive any new link since they arrive in the network with $q_i = 0$. This shows that in the BA model, the outdegree of nodes is their initial total degree from which the "proportional to degree" attachment rule applies. As noted by Newman (2010), Dorogovtsev et al.'s (2000) model is an extension of de Solla Price's (1976) model, who used $a = 1$. At this stage it is useful to remark that in this type of models, nodes compete to receive new edges. Hence to compute the *probability* that a given node receives a given edge, we always need to make sure that for each edge, the sum over all nodes of the probabilities that a node receives this edge is equal to one. Hence the attachment rule is often called the attachment kernel. Often the difficulty comes from computing the correct normalization factor. In the case

---

[1]A point emphasized by Dorogovtsev et al. (2000) is that in their version of the model, it does not matter whether the new edges come from the new node or from some others.

of (3.6), the total number of indegrees is equal to the number of outdegrees, which increases by m at each period and hence is equal to mt. Hence $\sum_{i=1}^{t}[q_i + a] = t(m + a)$. Thus (3.6) becomes

$$\frac{dq_i}{dt} = \frac{q_i + a}{t(1 + a/m)}. \tag{3.7}$$

Using this, the solution of (3.7) with initial condition $q_i(t_i) = 0$ is

$$q_i(t) = a \left[\left(\frac{t}{t_i} \right)^{\frac{1}{1+a/m}} - 1 \right], \tag{3.8}$$

and using the cumulative distribution as for the BA model in the beginning of this section, one obtains

$$p(q) \propto (q + a)^{-2 - \frac{a}{m}}, \tag{3.9}$$

where the symbol $\propto$ means "proportional to". The general form of the distribution (3.9), $p(k) = (k + r)^{-\gamma}$, will be called the "shifted power law" throughout the thesis (it appeared in Mandelbrot (1953) who insisted that the shift r was very helpful to obtain a better fit to the data when a "shoulder" appears on the left, that is, power law behavior is not observed for small k). It is close to a power law, in the sense that $k + a \to k$ as $k \to \infty$. However, for small k, the term a dominates and the function $p(k)$ decreases slower than a power law, creating a "shoulder" on the left of the distribution (the bins of the histogram for small k). From (3.9), note that we recover the BA model if we set $a = m$, and remember that $k_i = q_i + m$.

This model is very similar to models with mixed random and preferential attachment (reviewed in Jackson (2008)). Suppose that a node arrives and, out of its m links, a share α goes to nodes selected uniformly at random, and a share $1 - \alpha$ goes to nodes chosen with probability proportional to their indegree. Then

$$\frac{dq_i}{dt} = m \left[\alpha \frac{1}{t} + (1 - \alpha) \frac{q_i}{mt} \right] = \frac{q_i + \frac{\alpha m}{1 - \alpha}}{\frac{1}{1 - \alpha} t}, \tag{3.10}$$

showing that a mixed random and preferential attachment implies a "shifted" attachment kernel, and therefore a "shifted" power law

$$p(q) \propto \left(q + \frac{m\alpha}{1 - \alpha} \right)^{-1 - \frac{1}{1-\alpha}}. \tag{3.11}$$

For completeness, consider now that a node arrives and, out of its m links, a share α goes to nodes selected uniformly at random, and a share $1 - \alpha$ goes to nodes

chosen with probability proportional to their degree $k_i = q_i + m$. Then,

$$\frac{dq_i}{dt} = m \left[\alpha \frac{1}{t} + (1 - \alpha) \frac{q_i + m}{2mt} \right] = \frac{q_i + \frac{(1+\alpha)m}{1-\alpha}}{\frac{2}{1-\alpha}t}, \qquad (3.12)$$

leading to

$$p(q) \propto \left(q + \frac{(1+\alpha)m}{1-\alpha} \right)^{-1-\frac{2}{1-\alpha}}. \qquad (3.13)$$

To obtain the distribution of the total degrees, as in Jackson (2008) p.136, the change of variable $k = q + m$ gives the correct result, $p(k) \propto (k + \frac{2\alpha m}{1-\alpha})^{-1-2/(1-\alpha)}$. What is interesting in this exercise is that the specific way of microfounding a shifted attachment kernel by a mixed attachment mechanism does not change the shape of the distribution but changes the effects of the parameter m on the distribution. In Dorogovtsev et al.'s (2000) model (equations 3.7-3.9), m has no effect on the shift of the attachment kernel, and therefore no effect on the shift of the power law. However, m appears in the partition factor of the attachment kernel (the denominator of the differential equation), and therefore in the exponent of the power law. In the two mixed attachment models (equations 3.10-3.13), m does not appear in the denominator of the differential equation, and does not appear either in the exponent of the power law. To see that this pattern holds in general, it can be verified that

$$\frac{dq_i}{dt} = \frac{q_i + f(m)}{tg(m)}$$

gives

$$p(q) \propto (q + f(m))^{-1-g(m)}.$$

An important extension of the simple preferential attachment kernel is the non-linear kernel introduced by Krapivsky et al. (2000) in the original BA framework. At each period one node arrives and connects to m others, chosen with probability proportional to a certain power of their total degree:

$$\frac{dk_i}{dt} = m \frac{k_i^\alpha}{\sum_i k_i^\alpha}. \qquad (3.14)$$

When $\alpha = 1$, we recover the linear preferential attachment kernel. When $\alpha > 1$ (a superlinear attachment kernel), the vertices with a high degree get a disproportionate share of incoming edges, leading to a winner takes all situation (Krapivsky et al. 2000, Chung et al. 2003) in the very long run, but a scale free distribution for the non-leader nodes at finite time (Krapivsky & Krioukov 2008). When $\alpha < 1$, known as the sublinear attachment kernel case, we can argue that (Liu et al. 2002) $\sum_i k_i^\alpha = \mu t$, with $1 < \mu < 2m$ since $\sum_i k_i^0 = t < \sum_i k_i^\alpha < \sum_i k_i^1 = 2mt$. Using this,

equation 3.14 becomes

$$\frac{dk_i}{dt} = m\frac{k_i^\alpha}{\mu t}.$$ (3.15)

Integrating (3.15) with the initial condition $k_j(j) = m$ gives (Liu et al. 2002)

$$k_i(t) = \left(m^{1-\alpha} + \frac{(1-\alpha)m}{\mu}\ln\left(\frac{t}{t_i}\right)\right)^{1/(1-\alpha)},$$ (3.16)

which can be used with the usual method to obtain the stretched exponential form

$$p(k) = \frac{\mu}{m}k^{-\alpha}\exp\left\{-\frac{\mu}{m}\frac{k^{1-\alpha}-m^{1-\alpha}}{1-\alpha}\right\}.$$ (3.17)

The tail ($k \to \infty$) of this distribution is between the exponential ($\alpha = 0$) and a power law with exponent 3 when $\alpha = 1$ (which recovers Barabási & Albert (1999)). The non-linear attachment model is of particular interest in this thesis. In chapter 4, where I study a growing two-mode network, the micro-level assumptions (learning random unknown ideas of random friends) lead to a non linear attachment kernel of a different form than that studied by Krapivsky et al. (2000) and others. In my model, the attachment kernel has a logistic form, that is, proportional to $k(n-k)$ with n being the maximum degree, which in this case is population size (an idea cannot be known by more agents than there exist). Moreover, chapter 4 highlights that the derivations above *assume* that μ is a constant, whereas it is constant only for large time/network size. In my argument, μ testifies the self-organization of the network in the sense that the dynamics of the distribution depends on μ, and μ itself depends on the distribution. In fact, μ determines the dynamics and the distribution represents the structure. Their convergence to a steady-state shows that the feedback mechanism leads to a fixed point, which is *de facto* self-organized.

In chapter 4, the main method used is different, although the mean-field continuous approximation is also used for comparison purposes. The master equation method, which is widely used for deriving the equilibrium distribution of Markov chains, is more precise than the mean-field continuous approximations of this section. It works very well for linear and shifted attachment kernel, as exemplified in the original papers of Dorogovtsev et al. (2000) and Krapivsky et al. (2000). However, for a sublinear attachment kernel, it is not possible to obtain exact solutions of the master equation in terms of well-known functions (Iguchi & Yamada 2007). It turns out that the logistic attachment kernel of chapter 4, which is quadratic and not of a fractional power, is less problematic. Chapter 4 suggests that in general, a polynomial attachment kernel leads to a generalized hypergeometric structure of the master equation.

The next two sections introduce master equations in the context of some important models of growing *two-mode* networks.

3.2 Simon's (1955) model

More than half a century ago, Herbert Simon proposed a model of stochastic growth to explain the emergence of power law distributions. This model has been widely used since then, and even started a new life after the general increase of interest in networks. This model has a rich history and some intricacies have been clarified only recently. The purpose of this section is to provide a detailed review of Simon's model.

In general terms, consider a number of elements grouped into a number of entities. Let us think of elements (or items) as inhabitants and entities (or categories) as cities. To understand the static picture (the observed distribution), we have to understand the dynamic process which leads to it. Simon hypothesized that the system only grows. There are new inhabitants arriving, say one per period (time is "system time", not "clock time"). All that has to be chosen is if newborns create a new city, or to which city they go if they go to an existing one. Simon realized that if he assumes that existing cities get the new inhabitant with a chance proportional to the size of their population, the bins of the (normalized) histogram will ultimately stabilize. He recalls being aided by a metaphor, while thinking about the power law distribution of the number of times that a word is used in a text.

> If we think of a book as being created word by word, and if a word is added that has occured k times, the number of words occuring $k + 1$ times each will be increased by one, and the number of words occuring k times each decreased by one. For a steady state equilibrium, the rate at which words are created that had previously occured k times must be equal to the rate at which words are created that had previously occured $k - 1$ times. In this way, the "k" bin will be replenished as rapidly as it is depleted. At some point I began to visualize this as a cascade, with successive pools of water each maintained at a constant level by flow in from the pool above, and flow out from the pool next below. Working back from our answer – the distribution that we know describes the phenomena - it is not too hard to show that the equilibrium condition requires that the probability of creating a word that has already occured k times must be proportional to k. (Simon 1989)

The framework sketched above leads directly to the formulation of a master

equation,[2] and the equilibrium condition provides a direct way to find the steady state.

The model assumes that discrete items arrive to a growing population of entities. At each period exactly one item arrives. Simon formally states two assumptions

Assumption 3.2.1. *With probability b, the new item goes to a new entity which is thus created.*

Assumption 3.2.2. *If the new item does not create a new entity, it goes to an entity chosen with probability proportional to its size (number of items within this entity).*

The master equation should tell us the expected change in the number of entities of size k, denoted $P(k)$. This number increases by one when a new item goes to an entity of size $k - 1$, and decreases by one when a new item goes to an entity of size k. Suppose that the new item has to go to an existing entity (this happens with probability $1 - b$). The probability that the new item goes to an entity of size k is proportional to k, by assumption 3.2.2. Conditional on the new item not creating a new entity, it must be that exactly one entity is chosen. This entity is in exactly one class of size (say the kth bin). So it must hold that the sum of the probabilities over all possible classes of size (i.e. over k) is one. So the probability (conditional on creating a new entity) that the new item goes to an entity of size k is k divided by the total size of all other competing entities. This total is simply the total number of items in the system, t.

The expected number of entities of size k in $(t+1)$ is equal to number of entities of size k in t, plus the number of entities of size $k - 1$ in t times their probability of getting the new item, minus the number of entities of size k in t times their probability of getting the new item. Putting this together,

$$E[P_{t+1}(k)] - P_t(k) = (1 - b)\left(P_t(k - 1)\frac{k - 1}{t} - P_t(k)\frac{k}{t}\right). \qquad (3.18)$$

The system always grows so the vector $P(k)$, the total number of entities of a certain size, can never stabilize. We need to study the *share* of entities of a certain size, which is a probability mass function defined as $p_t(k) = P_t(k)/w_t$, where the total number of entities w_t is a stochastic variable. If $\Pi_w(t)$ is the probability that in t there are exactly w entities in the system,

$$\Pi_w(t) = \binom{t - 1}{w - 1}b^{w-1}(1 - b)^{w-t}, \qquad (3.19)$$

---

[2]The term master equation is used in this context for an equation describing the expected changes in the heights of the bins of a histogram. It is used more generally in the context of Markov chains, for describing the change in the probabilities that a system be in certain state.

from which

$$E[w_t] = E[kP_t(k)] = E[\Pi_w(t)] = \sum_{j=1}^{t} j\Pi_j(t) = 1 - b + bt, \quad (3.20)$$

where the expectation is over many realizations of the system. In general, $E[P_t(k)/w_t] \neq E[P_t(k)]/E[w_t])$. But from equation 3.20, we see that $E[w_t]$ tends to bt when $t \to \infty$. Simon therefore uses $w_t = bt$. With this, taking expectations of equation 3.18 gives

$$(t+1)p_{t+1}(k) - tp_t(k) = (1-b)[(k-1)p_t(k-1) - kp_t(k)].$$

Assuming that a stationary distribution exists for $t \to \infty$, so that $p_{t+1}(k) = p_t(k) = p(k)$, one obtains

$$p(k) = (1-b)\left[(k-1)p(k-1) - kp(k)\right]. \quad (3.21)$$

So far, only existing entities have been considered. Hence, the equations are not valid to describe the evolution of the number of entities of size 0, since they do not exist, or the number of entities of size 1, since it is affected by the creation of new entities. For $k = 0$, we know that $p(0) = 0$ since each entity enters with a degree equals to 1. For $k = 1$, there are no inflows of entities from the $(k-1)^{th}$ bin. Instead, there is a fixed inflow of new entities, at a rate b (assumption 3.2.1). The master equation for $k = 1$ therefore reads

$$P_{t+1}(1) = P_t(1) + b - (1-b)P_t(1)\frac{1}{t}.$$

Using the same simplification as before gives

$$p(1) = 1 - (1-b)p(1). \quad (3.22)$$

Hence the share of entities of size one, $p(1)$, can be found directly by solving equation 3.22

$$p(1) = \frac{1}{2-b}. \quad (3.23)$$

To solve equation 3.21, it is convenient to introduce $\hat{b} = \frac{1}{1-b}$, so that it can be rewritten as

$$p(k) = \frac{k-1}{\hat{b}+k}p(k-1), \quad (3.24)$$

which implies

$$p(2) = p(1) \ \frac{2-1}{\hat{b}+2}$$

$$p(3) = p(1) \ \frac{2-1}{\hat{b}+2} \frac{3-1}{\hat{b}+3}$$

$$p(k) = p(1) \ \frac{(k-1)(k-2)...1}{(\hat{b}+k)(\hat{b}+k-1)...(\hat{b}+2)}. \tag{3.25}$$

The numerator is $(k-1)! = \Gamma(k)$, where Γ is the Gamma function. The denominator should be seen as a ratio of Gamma functions $(\hat{b}+k)(\hat{b}+k-1)...(\hat{b}+2) = \Gamma(\hat{b}+k+1)/\Gamma(\hat{b}+2)$. Hence,

$$p(k) = p(1) \ \frac{\Gamma(\hat{b}+2)\Gamma(k)}{\Gamma(k+\hat{b}+1)} = p(1)(\hat{b}+1) \ \frac{\Gamma(\hat{b}+1)\Gamma(k)}{\Gamma(k+\hat{b}+1)}. \tag{3.26}$$

Using the definition of the beta function $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$,

$$p(k) = p(1)(\hat{b}+1) \ B(k, 1+\hat{b}). \tag{3.27}$$

Substituting (3.23) into (3.27), one obtains the Yule-Simon distribution

$$p(k) = \hat{b} \ B(k, 1+\hat{b}). \tag{3.28}$$

Another way to solve (3.24) is to remark that its series is a particular instance of Gauss hypergeometric series. Starting again from (3.25),

$$\sum_{k=1}^{\infty} p_k = p(1) + p(1)\frac{1}{\hat{b}+2} + p(1)\frac{1}{\hat{b}+2}\frac{2}{\hat{b}+3} + \cdots + p(1)\frac{1 \ 2 \ ...k}{(\hat{b}+2)(\hat{b}+3)...(\hat{b}+k+1)} + \cdots$$

$$= p(1)\left(1 + \frac{1}{\hat{b}+2} + \frac{1}{\hat{b}+2}\frac{2}{\hat{b}+3} + \cdots + \frac{1 \ 2 \ ...k}{(\hat{b}+2)(\hat{b}+3)...(\hat{b}+k+1)} + \cdots\right)$$

$$\sum_{k=1}^{\infty} p_k = p(1)\left(1 + \sum_{K=1}^{\infty}\prod_{i=0}^{K-1}\frac{i+1}{i+\hat{b}+2}\right). \tag{3.29}$$

By mere definition of Gauss hypergeometric function,

$$_2F_1\left[\{a_1, a_2\}, \{b_1\}; z\right] := \sum_{s=0}^{\infty}\frac{\Gamma(a_1+s)\Gamma(a_2+s)\Gamma(b_1)}{\Gamma(a_1)\Gamma(a_2)\Gamma(b_1+s)\Gamma(s+1)}z^s, \tag{3.30}$$

it can be seen that (3.29) is

$$\sum_{k=1}^{\infty} p_k = p(1)_2F_1(1, 1, \hat{b} + 2, 1), \tag{3.31}$$

and therefore, identifying the terms of the sum gives equation 3.26, and so the Yule-Simon distribution. This connection between the Yule-Simon distribution and the Gauss hypergeometric function is useful because one can use Gauss hypergeometric summation theorem,

$$_2F_1(a, b; c; 1) = \frac{\Gamma(c)\Gamma(c - a - b)}{\Gamma(c - a)\Gamma(c - b)}, \tag{3.32}$$

with $a = 1$, $b = 1$ and $c = \hat{b} + 2$ to evaluate (3.31) and check that it is equal to one.[3]

In fact, in the study of growing network models and other Polya urns, one often resorts to the use of generating functions, which allow to derive a differential equation for the series of the distribution. It is also known that hypergeometric functions satisfy a certain differential equation. This suggests a link between the differential equation obtained by using generating functions and the one satisfied by the Gauss hypergeometric function, which is the Gauss hypergeometric differential equation:

$$z(1 - z)\frac{d^2 f(z)}{dz^2} + (c - (a + b + 1)z)\frac{df(z)}{dz} - abf(z) = 0.$$

To uncover this link, define the generating function

$$g(z) = \sum_{k=1}^{\infty} p(k)z^{k-1}.$$

It implies $zg' + g = \sum_{k=1}^{\infty} kp(k)z^{k-1}$ and $z(zg' + g) = \sum_{k=1}^{\infty}(k - 1)p(k - 1)z^{k-1}$. Summing up the master equation 3.21 from $k = 1$ to ∞ and using these definitions, one gets

$$\hat{b}g + (zg' + g) - z(zg' + g) = 0,$$

that is

$$(\hat{b} + 1 - z)g + (z - z^2)g' = 0.$$

---

[3]Ijiri & Simon (1977) use the integral representation of the beta function to prove this result. The original paper of Simon does not contain this result (perhaps due to the typo in equation 2.13 p. 429), but gives arguments that the series converge when the power law exponent is sufficiently high. Garibaldi & Scalas (2010) p.268 prove normalization by showing that the partial sum of the terms tends to one as the number of terms accounted for goes to infinity. For the case of the Waring distribution, Peterson et al. (2010) in a footnote on p. 16024, use a formula for infinite sums of factorials. Fenner et al. (2005) used Gauss theorem.

Following Iguchi & Yamada's (2007) procedure for the BA model, take derivatives with respect to z and simplify to get

$$z(1 - z)g'' + (\hat{b} + 2 - 3z)g' - g = 0,$$

which is Gauss hypergeometric differential equation with parameters $a = 1$, $b = 1$, $c = \hat{b} + 2$, as it should. As far I am aware of, the derivation of Gauss hypergeometric differential equation for Simon's model is original, although (3.31) and the relevance of Gauss Hypergeometric Theorem had previously been noticed (Fenner et al. 2005). They are not necessary here, but they highlight a more general theory behind Simon's model. In chapter 4, the master equation will be non-linear due to the logistic diffusion of ideas. To prove that the obtained distribution is properly normalized, a more general hypergeometric theorem (Pfaff-Saalschütz) will be used. This highlights that not only the distribution obtained in chapter 4 generalizes the Yule-Simon, but there exists a whole class of under-studied "Saalschützian" Hypergeometric Distributions which generalizes the class of "Gaussian" Hypergeometric Distribution (Gaussian Hypergeometric Distribution include the Yule-Simon and many well known others, see Johnson et al. (2005)).

Finally, note that in this section the existence of a steady-state distribution has been *assumed*. Several papers have studied convergence using different methods. Kullmann & Kertész (2001) have given complete solutions for the time-dependent marginal distributions, that is, the probability that a randomly chosen entity in a randomly chosen realization of the system exhibits degree k. From this expression, they derived the time-dependent distribution. Levene et al. (2002) studied perturbed steady-state equations. Hou et al. (2009) use Markov chains to analyze the evolution of individual nodes, and show that at the level of the bins of the histogram, a theorem on the convergence of series (Stolz-Cesàro) can be invoked.

3.3   Self-organizing two-mode networks

Simon's model can be seen as a self-organizing two-mode network. It has two types of nodes (say items and categories, authors and papers, inhabitants and cities...), and it is self-organizing in the sense that there is a feedback loop (cumulative advantage) which nevertheless leads to a clear pattern of organization (a power law size distribution). However, Simon's model is too simple for elaborated empirical settings. For instance, it reproduces Lotka's law (a power law distribution of the number of papers per author), but it makes the counter-factual assumption that papers are always single-authored. This makes the model quite unable to go along the well documented rise of collaborations in science. Ramasco et al. (2004) pro-

posed a series of modifications to Simon's model to study the self-organization of collaboration networks.

Ramasco et al. (2004) assume that at each period, $\bar{m}$ new authors form with $\bar{n} - \bar{m}$ existing authors a new team of size $\bar{n}$ to write 1 paper. Existing authors are chosen with probability proportional to their existing publication record. This is an extension of Simon's model to papers with multiple authorship. Ramasco et al. (2004) showed that the number of papers per author keeps the form of the Yule-Simon distribution

$$p(k) = \frac{\bar{n}}{\bar{n} - \bar{m}} \mathrm{B}\left(2 + \frac{\bar{m}}{\bar{n} - \bar{m}}, k\right). \tag{3.33}$$

A related model was proposed by Peruani et al. (2007). In their model, one set of nodes is an alphabet, that is, a set of elements that can be combined. The other set represents the combinations. Every period a new combination is created, and has m edges towards the elements of the alphabet it uses. So, as in Simon's and Ramasco et al., one side of the network is growing of one node per period. But in this case, the other set of nodes – the alphabet – is fixed, equal to N. They study the degree distribution of the fixed set of nodes, assuming a shifted attachment kernel, which they choose to write $A(k,t) \propto \gamma k + 1$. This model does not have a steady state, but they found an insightful approximation for the time-dependent distribution in the long time limit

$$p(k,t) \propto (k/t)^{1/\gamma - 1}(1 - k/t)^{\frac{N}{\gamma m} - 1/\gamma - 1}.$$

Depending on the strength of preferential attachment, they show that this distribution varies from binomial ($\gamma \to 0$) to U-shaped ($\gamma > N/m - 1$).

This behavior was found in another model of a self-organizing bipartite graph due to Evans (2007) (see also Evans & Plato (2008)). This model has a fixed number of nodes (E agents and N artefacts) and edges. At each period, only one end-of-edge is (potentially) modified, always on the same side. In other words, each agent is connected to exactly one artefact, and at each period, one edge is detached from an artefact, and attached somewhere (possibly back to the same artefact). The departure node is chosen with preferential probability, and the arrival node is chosen with a shifted attachment rule. As emphasized in Evans (2007), it is important to account for the case in which the vertex from which the edge is removed is also the one to which the edge is attached. The equilibrium distribution of Evans's (2007) model is found by showing that the generating function satisfies Gauss Hypergeometric differential equation. The equilibrium distribution is thus expressed in terms of Gamma functions (see Evans (2007) for the exact expression).

For large k, and if the attachment kernel is sufficiently random (not preferential), this can be approximated either by a distribution of the beta form

$$p(k) \propto k^{-\gamma} \left(1 - \frac{k}{E} \right)^{E\bar{\xi}},$$

or as a power law with exponential cut-off

$$p(k) \propto k^{-\gamma} e^{-\xi k}.$$

However, when the attachment is sufficiently preferential, Evans (2007) observes that an upward curvature appears, as in Peruani et al. (2007).

These models are particularly relevant for this thesis as they showed practically how to analyze self-organizing two-mode networks, and are the closest to the model of chapter 4. Other contributions include Peltomäki & Alava (2006), who studied explicitly degree correlations, and Hébert-Dufresne et al. (2011) who propose a framework encompassing two-mode and one-mode growing network models. Criado et al. (2012) proposed a model based on hypergraph theory, where a whole subgraph is added at each period, which inspired the model of the next section, where I propose to generalize Simon's and Ramasco et al's by adding a complete new two-mode subgraph at each period.

3.4   Growth by addition of full subgraphs

In this section, I develop a simple model that encompasses Simon's and Rasmaco et al's models. In Simon's and Ramasco et al's models, there is one more new paper per period, and some new authors (In Simon's, a new author arrives with a given probability, which corresponds to having a fraction of a new author per period). Neither Simon nor Ramasco et al. attempt to answer the question of what happens if several new papers and several new authors arrive at each period. I address this question explicitly here assuming that all the new authors are co-authoring all the new papers. The idea is to add a given full subgraph at each time step. Some edges are going out and are used to connect to the existing graph. These connections are made using preferential attachment kernels. In particular, preferential attachment of new nodes of type X to existing nodes of type Y (i.e. with probability proportional to the degrees of Y-nodes) generalizes the famous model of Simon (1955) and the "self-organization of collaboration networks" model of Ramasco et al. (2004), which are recovered as particular cases. I show that there exists a relationship between the power law exponent of the degree distribution of

each set of nodes and the number of incoming edges/new nodes of both sets of nodes. As in section 3.1, the main insight is that the power law exponent depends on the relative weights of random and preferential attachment. Since attachment is always preferential (by assumption here), the "random" (non-preferential) part of attachment comes from the initial degree of nodes. The model is shortly discussed in its original context of author-paper networks. If we think of projects as sets of people and papers all connected and arriving simultaneously, the results imply that Lotka's law is influenced by project's size.

A two-mode network has two degree distributions. However, the model is symmetric so the equations presented are valid under relabelling of the two partitions. Let the two partitions be labeled X and Y. We are looking for the probability mass function for the degrees of Y-nodes, denoted $p^Y(k)$. Consider an existing graph G. At each period, a new subgraph, called g, is added to G. This new subgraph has three parts:

- A complete two-mode subgraph denoted $\bar{g}$. It has n^X X-nodes, n^Y Y-nodes, and $(n^X \cdot n^Y)$ edges,

- A number of edges from each X-node in g to some Y-nodes in G: m_o^Y,

- A number of edges from each Y-node in g to some X-nodes in G: m_o^X.

Consider one of the new edges going from new X-nodes to old Y-nodes. Then $A_t^Y(k)$ denotes the probability that a Y-node having degree k at time t receives that particular link. Here I consider preferential attachment, $A_t^Y(k) \propto k$.

An initial condition that is simple and will deliver nice equations is as follows. At time t_0 there is nothing. We start at time $t = 1$ with g, although some edges are only stubs. They are not connected to anything, but we count them both in the degrees of the "emitting" nodes and in the total number of degrees[4]. From $t = 2$ onwards we continue as described in the algorithm. With this convention, the total number of X-nodes at time t is simply $w_t^X = n^X t$. Let us also introduce the total number of edges in g, $n^E = n^X n^Y + n^X m_o^Y + n^Y m_o^X$. Finally, it is useful to define $v = n^X m_o^Y$. This is the exact number of new edges which connect to existing Y-nodes. These ones are the key of the model, since they are those whose attachment is governed by the attachment kernel. In the sequel, I assume that the new subgraphs are relatively small. Multiple edges are allowed to avoid complications, but there tends to be fewer and fewer of them as time goes to infinity.

---

[4]This is just to get rid of constants that add pedagogical complications. It does not change the nature of the results, since these stubs become anyway negligible as the system grows.

Using the above assumptions and approximations, the master equation for the number of Y-nodes with degree k is:

$$P_{t+1}^Y(k) - P_t^Y(k) = v\left(P_t^Y(k-1)A_t^Y(k-1) - P_t^Y(k)A_t^Y(k)\right),\qquad(3.34)$$

and its initial condition

$$P_{t+1}^Y(q) - P_t^Y(q) = n^Y - vP_t^Y(q)A_t^Y(q),\qquad(3.35)$$

where $q = n_X + m_o^X$ is simply the degree of an incoming Y-node. The classical "rich gets richer" attachment kernel is chosen,

$$A_t^Y(k) = \frac{k}{\sum_{j=1}^{w_t^Y} k_j} = \frac{k}{n^E\, t},\qquad(3.36)$$

where n_E is the number of edges in g. Second, the probability mass function $p(k)$ giving the share of Y-nodes having degree k is the number of Y-nodes with degree k $(P(k))$ divided by the total number of Y-nodes $(w_t^Y = tn^Y)$,

$$P_t^Y(k) = w_t^Y p_t^Y(k) = n^Y\, t\, p^Y(k).\qquad(3.37)$$

Inserting equations 3.36 and 3.37 into 3.34 and 3.35, and assuming a steady state such that $p_{t+1}(k) = p_t(k) = p(k)$, one obtains

$$p^Y(k) = \frac{k-1}{(n^E/v) + k}\, p^Y(k-1),\qquad(3.38)$$

and

$$p^Y(q) = \frac{1}{1 + (v/n^E)q}.\qquad(3.39)$$

Now, solving the recurrence equation 3.38 with initial condition 3.39 nicely simplifies to

$$p^Y(k) = \frac{B(\alpha + 2, k)}{B(\alpha + 1, q)},\qquad(3.40)$$

where $B(,)$ is the beta function and

$$\alpha = \frac{n^E}{v} - 1 = \frac{n^Y m_0^X + n^X n^Y}{n^X m_0^Y}.\qquad(3.41)$$

can be seen as the number of new edges going to new Y-nodes relative to the number of new edges going to old Y-nodes. Notice that the sum of the three elements in the RHS of (3.41) gives n^E, the total number of new edges. So the exponent of the power law depends on how the total number of new edges is split up into its

	n^X	n^Y	m_o^Y	m_o^X	$\gamma - 2$
Simon (1955)	1	b	$1 - b$	0	$\frac{b}{1-b}$
RDP (2004)	1	m	$\bar{n}$-$\bar{m}$	0	$\frac{\bar{m}}{\bar{n}-\bar{m}}$
Full subgraphs	n^X	n^Y	m_o^Y	m_o^X	$\frac{n^Y m_o^X + n^X n^Y}{n^X m_o^Y}$

Table 3.1: The values are given assuming that the partition studied is p^Y. The last column gives the power law exponent $B(k, \gamma) \sim k^{-\gamma}$ from equations 3.28, 3.33 and 3.40.

.

three parts: the number of edges of the full subgraphs, i.e. the edges going from new X-nodes to new Y-nodes $(n^X n^Y)$, the number of edges going from new Y-nodes to existing X-nodes $(n^Y m_0^X)$, and the number of edges going from new X-nodes to existing Y-nodes $(v = n^X m_0^Y)$. Hence, the numerator is related to the degrees of the *new* Y-nodes, whereas the denominator is related to the new degrees for the *old* Y-nodes. Thus, we find again Simon's insight: the exponent for the degree distribution of Y-nodes must depend on the probability (in this case share) of new (vs. existing) Y-nodes. This result comes from the fact that the edges of newborn Y-nodes are given uniformly (each node has the same chances to get those edges; in fact they all have exactly the same number of edges initially), whereas edges obtained afterwards are given preferentially.

The model admits two interesting particular cases described below and summarized in table 3.1. Ramasco et al. (2004) assume that at each period, $\bar{m}$ new authors form with $\bar{n} - \bar{m}$ existing authors a new team of size $\bar{n}$ to write 1 paper (section 3.3). We want to derive the degree distribution of authors, $p^Y(k)$, that is, the probability that a randomly chosen author has written k papers. The number of new X-nodes is 1, the number of new Y-nodes is $\bar{m}$. The number of new edges that goes from a new author to an old paper is zero $(m_o^X = 0)$,[5] and the number of new edges that go from a new paper to an old author is $m_o^Y = \bar{n} - \bar{m}$. From equation 3.41, we can compute $\alpha = \frac{\bar{m}}{\bar{n}-\bar{m}}$, giving an exponent in agreement with their results (equation 3.33).

As a second example, consider Simon's model (section 3.2). Here one X-node arrives with one edge which attaches to one new Y-node with probability b, and to an old Y-node with probability $(1 - b)$. From the master equation point of view, we can think of it as a new X-node arriving with b new Y-nodes and $(1 - b)$ old Y-nodes attached. We have $n_X = 1$, $n_Y = b$, $m_o^Y = 1 - b$ and $m_o^Y = 0$, giving again the correct formula 3.28. Of course, it should be noted that in Simon's model b is a probability, and here m_o^Y is treated as an integer. But from the point of view of a rate equation for the height of the bins of the degree distribution, it does not matter whether we have integer or rational numbers – a point already demonstrated

---

[5]This possibility would imply imitation or diffusion, and is discussed in chapters 4 and 5

in practice by Ramasco et al. (2004).

An important limitation of this model is that existing authors are not full members of the projects. Only new authors are linked to all new papers, but the existing authors of new papers are chosen by one paper at a time.

3.5   Conclusion

This chapter presented a number of recent models of self-organizing multi-mode networks. Two important models in the context of knowledge systems, Simon (1955) and Ramasco et al. (2004), were shown to be particular cases of a two-mode network growing by addition of full subgraphs, which clarifies the origins of the power law exponent in these models.

A classical critique of economists to this type of work is the lack of clear (axiomatic) behavioral foundations. In the next chapter, I start from justifiable premises about behavior to *derive* the attachment kernel of a growing two-mode network of agents and ideas.

Chapter 4

Self-organization of knowledge economies

Abstract

Suppose that homogenous agents fully consume their time to invent new ideas and learn ideas from their friends. If the social network is complete and agents pick friends and ideas of friends uniformly at random, the distribution of ideas' popularity is an extension of the Yule-Simon distribution. It has a power-law tail, with an upward or downward curvature. For infinite population it converges to the Yule-Simon distribution. The power law is steeper when innovation is high. Diffusion follows S-shaped curves.

4.1   Introduction

The importance of knowledge in explaining economic outcomes has been widely documented. At the individual level, educational training and skills determine income (Schultz 1961) and capabilities (Sen 1999). At the firm level, innovation is the source of competitive advantage and profits (Schumpeter 1934). At the country level, technological change explains most of GDP growth (Solow 1957).

To understand the process of economic development, one should therefore study the generation and diffusion of ideas. The literature on endogenous growth has significantly clarified the mechanisms through which knowledge can lead to GDP growth (Lucas 1988, Romer 1990), but less efforts have been devoted to the study of the detailed distribution of ideas in simple, decentralized "knowledge economies" in which agents create and exchange ideas. Some patterns are more likely or efficient than others (Cowan & Jonard 2004). For economists, it is crucial to have expectations about the structure of who knows what (the distribution of ideas). For

instance, since production relies on knowledge, the structure of who knows what influences the structure of who produces what (product differentiation and countries' specialization). Moreover, since knowledge is (mostly) a public good, and is (mostly) cumulative in nature, the structure of who knows what determines and is determined by the rate and direction of inventive activity. Therefore, long-run economic progress depends intimately on the detailed organization of knowledge systems.

This chapter analyzes the structure of who knows what by deriving the distribution of ideas' popularity in a simple model based on the assumption that attention is allocated between innovation and imitation. The objectives are first, to find conditions under which a stable system can be characterized, and second, to characterize the resulting organization. In other words, this chapter studies the self-organization of knowledge economies: when a collection of agents produce and consume knowledge, can we expect a certain form of stability in the distribution of knowledge? and if so, which is this stable form? I find that if the trade-off between innovation and imitation is constant, then a stable distribution of ideas' popularity emerges, in spite of the disturbing force of innovation being at play, with new ideas arriving regularly. Moreover, even though which ideas diffuse and which agents are chosen to receive and diffuse knowledge are stochastic events, self-organization produces a certain stability in the average overlap among agents' ideas' portfolios, and hence in the distribution of ideas' popularity. Self-organization can be understood at the mean-field level, where there exists a fixed point, self-consistency equation from which one can derive a steady-state that is unique. In other words, when agents create new ideas and learn random ideas of random friends, after some time the structure of who knows what will be such that the diffusion process is compatible with that same structure, even though it is growing due to innovation. Hence it is a stable, self-organized knowledge economy.

The distribution of ideas' popularity is, roughly speaking, a power law, due to the fact that learning random ideas of random friends produces cumulative advantage (or self-reinforcing dynamics) in ideas' diffusion: the more an idea is known, the higher the chances that it is found at random in a random friend. However, since population is bounded, which ensures logistic (S-shaped) diffusion, the power law has finite support (ideas are known by at most the number of agents in the population). When the social network is complete, this finite support power law is characterized precisely, as a discrete distribution which is a particular case of a Generalized Hypergeometric Distribution, and an extension of the Yule-Simon distribution. Changing the social network can change the distribution of ideas' popularity to some extent, and this is investigated mostly using simulations.

The relationship between the special Pfaff-Saalschützian Generalized Hyergeo-metric distribution derived here and the Yule-Simon distribution follows from the fact that the proposed model can be seen as an extension of Simon's (1955) model. In Simon's original model, there are agents and ideas. At each period, a new idea arrives. With some fixed probability b, it goes to a new agent (created simultane-ously). Otherwise, it goes to an agent chosen with probability proportional to the number of ideas that he holds. This process leads to a steady-state distribution of the number of ideas per agent which has power law tail, and is called the Yule-Simon distribution. For instance, among many other fields of application, Simon fitted his distribution using scientific authors and their papers. My starting point is that diffusion is missing. Scientific papers, like technologies and social norms, diffuse through the population. For clarity let me abstract from agent heterogeneity, and consider a fixed number of agents. I still want to have a growing number of ideas, consistent with reality, but also wish to allow agents to learn ideas of/from others. Since I contend that attention is limited, I assume that at each period, a randomly chosen agent chooses *either* to innovate, *or* to learn an existing idea. The agent then gets a new edge in the two-mode network of agents and ideas, a (bipartite) network where an edge between agent i and idea j means that "i knows j"'. The other side of the new edge is either an existing idea or a new one (figure 4.1). As described, the process is close to Simon's, but with one *fixed* set of nodes. This is important because the finiteness of population is necessary for diffusion to be S-shaped. Simon's master equation for the degree distribution should be modified, using a quadratic instead of linear attachment kernel. The resulting distribution is an extension of the Yule-Simon distribution, and resembles the beta distribution. It converges to the Yule-Simon when the population is infinite. The Yule-Simon distribution has one parameter, which depends on the relative rates of innovation and imitation.

The chapter is organized as follows. The next section discusses related litera-ture. Section 4.3 presents the model and clarifies key mathematical relationships in this setup. Section 4.4 gives the main results. Section 4.5 provides some results for two key generalizations (with a sparse social network, and with differentiated pro-ductivity of the time spent on imitation or innovation). The last section concludes.

4.2 Related literature

Cohen & Levinthal (1989) argued that R&D activities allow firms to absorb knowl-edge spillovers from their environment, reinforcing innovation capabilities. This paper is about the global organization of knowledge systems resulting from the

allocation of time between "true" (new-to-the-world) innovation and learning/ imitation/ diffusion. There is an extensive literature on diffusion (Geroski 2000), but it generally takes the new technology (idea, product, etc.) as pre-existing, and is concerned only with adoption, elaborating the mechanisms behind diffusion and looking for their idiosyncratic traces in empirical data (Young 2009). In the information age, much more information is available online, and we could think that individuals and firms learn from a database, instead of from their friends. Why would neighborhood effects in learning be so important then? One reason is that knowledge is tacit, situated, localized or embedded. This stickiness of knowledge implies that it can diffuse only, or preferably, face-to-face (Breschi & Lissoni 2009). Relatedly, social embeddedness channels awareness of ideas: one may learn new knowledge from a book or online after having been referred to it (by a peer). An important consequence of word-of-mouth interaction is that the diffusion pattern is likely to be S-Shaped, in agreement with the literature on diffusion (Mansfield 1961). In fact, learning from others naturally introduces increasing returns in ideas' diffusion due to the fact that well known ideas have more chances to diffuse, because they have more carriers. In the model below, as in the literature, this exponential growth is constrained by the population size in such a way that the diffusion is logistic. Logistic diffusion is well established theoretically and empirically, which leads to the two following questions: (i) What happens when there are *many* ideas competing for attention? (ii) What happens when there is continuous arrival of *new* ideas?

A way to characterize a system in which ideas are created and diffuse is by keeping track of the distribution of of ideas' popularity.[1] In the language of networks, this is the degree distribution of the "ideas" set of a two-mode network of agents and ideas. I assume one fixed set of nodes (the number of agents does not change) and one growing set of nodes (the number of ideas increases without bound). This framework allows to keep track of who knows what in a very detailed way, and provides a bridge between social network models (a one-mode network of agents) and epistemic network models (a one-mode network of ideas). Such a representation of the co-evolutionary dynamics of social and knowledge networks has been used in empirical studies (Roth & Cointet 2010) and simulation modelling (Börner et al. 2004). Cowan & Jonard (2009) study a closely related system, where firms form

---

[1]For models of knowledge growth and diffusion which do not involve networks, see e.g. Jovanovic & Rob (1989), König et al. (2012) and Lucas Jr & Moll (2014). The model presented here is complementary, because these models are more elaborated in terms of agent's choice and economic observables (e.g. GDP or productivity), but my model is richer in terms of the underlying combinatorial structure. For instance, since ideas are discrete in the model below, two agents with the same number of ideas can imitate ideas of each other, whereas two agents with the same productivity level cannot learn from each other in e.g. Lucas Jr & Moll (2014).

an alliance network based on knowledge matching. In their model, firms hold ideas and form pairwise alliances with other firms in order to innovate. Partner choice is based on knowledge overlap: too much overlap would mean that partners have few things to learn from each other; too little overlap may hinder mutual understanding. Their model reproduces several empirical facts of R&D networks, such as small world properties and skewed degree distribution.

The model developed here is also closely related to models of network growth based on copying (Vázquez 2003). For instance, in Jackson & Rogers's (2007) model for social networks, newborn agents choose to link to random existing agents ("random meetings"), and to random neighbors of their random meetings ("search"). Here, cumulative advantage comes from search meetings, because the more friends an agent has, the higher the chances he has to be found through a friend. Likewise, in a two-mode network, search can generate cumulative advantage and, ultimately, fat tail distribution of popularity. This was clearly demonstrated by Evans & Plato (2008), who consider a fixed set of agents and a fixed set of artefacts. Agents are linked to one and only one artefact, and, when they are chosen, connect to another artefact by imitating a friend. Their model is a two-mode network with both sets of nodes fixed, and a rewiring process. Actors are linked to one and only one artefact, and the distribution of artefacts popularity is studied. Their model applies for instance in anthropology where one is interested in the transmission of cultural artefacts. The model proposed below also applies to this context, but assumes that new artefacts appear over time, and that actors accumulate artefacts over time.

Another closely related model was studied by Ramasco et al. (2004). As Simon, they consider only the production of ideas (papers) but the number of agents is allowed to grow and papers are co-authored. Their work focused on reproducing the empirical data on the "co-authorship" network. Assuming that authors are chosen for new authorship with probability proportional to the number of their previously authored papers, Ramasco et al. (2004) derive the Yule-Simon distribution (with modified parameters) for the distribution of the number of papers authored by an author, and a shifted power law for the degree distribution of the co-authorship network. There have been other studies of two-mode or multi-mode networks in which all sets of nodes are growing. As the model presented here, Peruani et al. (2007) studied a model in which only one set of nodes is growing, but they analyze the degree distribution of the non-growing side, while the focus here is on the growing side. Beguerisse-Diaz et al. (2010) studied a system in which users rate videos. Liu et al. (2011) study a social tagging system, which can be seen as a three-mode network (users, resources tagged, and tags). Zeng et al. (2012) show that certain recommender systems produce more unequal popularity distribution

than others.

The model proposed below contributes to the literature on "self-organizing" networks by providing a detailed analysis of the artefact degree distribution under the assumption of a non-growing population of actors and assuming a specific one-mode network for agents' interactions. Technically, the most noticeable feature of the model presented below is that the probability for a given idea to diffuse at time t (the attachment kernel) is a *quadratic* function of its popularity at time t. This gives rise to a combinatorial interpretation of the partition factor of this attachment kernel. In the classical growing network model with sub-linear attachment kernel (Krapivsky et al. 2000), the value of the partition factor of the attachment kernel cannot be solved for in closed-form, and is computed numerically. In the model below, such a solution may exist but it is hard to find as it involves solving a polynomial of order n (number of agents).

More generally, the model presented here relates to a larger literature on re-inforcing processes in economics (Arthur 1989, Kirman 1993). Following Simon's own applications of his model, notably to the size distribution of firms (Ijiri & Simon 1977), there has been a large literature impossible to review here. In an influential contribution, de Solla Price (1976) applied Simon's process to explain the power law distribution observed for the in-degree of citation networks. He assumed that existing papers are cited with probability proportional the number of citations that they have already received. This assumption can be microfounded, by assuming that papers are found by searching through the bibliography of other papers (Vázquez 2003). The model below allows for an alternative microfoundation of citation networks. In the next chapter, using the model described below, and assuming that (an infinite number of) agents cite papers chosen uniformly at random among the papers that they have previously learned or written, the predicted citation distribution is a shifted power law.

4.3   The model

Before turning to the technical presentation of the fully fledged model, it is useful to present the main assumptions and results as follows:

Assumption 4.3.1 (Knowledge growth and innovation). *Knowledge is a set of discrete ideas. This set is expanding because new ideas are invented over time.*

Assumption 4.3.2 (Social embeddedness and diffusion). *Agents imitate ideas of their friends. More precisely, agents choose uniformly at random (u.a.r.) an (unknown) idea of a friend chosen u.a.r.*

Assumption 4.3.3 (Limited attention and innovation/imitation trade-off)**.** *Homogenous agents supply inelastically a fixed amount of attention to obtain ideas. Because some ideas must be invented (assumption 4.3.1), and some must be imitated (assumption 4.3.2), attention is split between these two activities. I assume that this split is the same for all agents and is constant over time.*[2]

Result 4.3.1. *Social embeddedness creates cumulative advantage for ideas' diffusion, that is, if diffusion was unbounded, ideas would diffuse at a rate proportional to their current popularity.*[3] *However, diffusion is constrained by the population size, as in logistic diffusion models. Thus, any idea's diffusion is S-shaped.*

Result 4.3.2. *This logistic diffusion of sequentially created ideas gives rise to a steady-state distribution of ideas' popularity which is close to a power law but with an upward or downward curvature in the tail. This curvature disappears when $n \to \infty$ and the distribution is the Yule-Simon distribution. A higher share of attention devoted to innovation (respectively, imitation) generates a steeper (flatter) power law.*

Since the model is stochastic, and because the boundedness of the population introduces a non linearity, the derivations of the steady-state distribution are rather involved and can obscure the gist of the argument. Hence, before going into the details of the stochastic model, let me first present a simplified version of the model that allows to see why learning ideas of friends under an innovation/imitation trade-off gives rise to a power law distribution of ideas' popularity. The derivations are a two-mode version of Barabási & Albert's (1999) original procedure. Assume an infinite population and deterministic diffusion. Once an idea is invented, it diffuses. Since agents learn ideas of their friends, the more carriers an idea has, the more chances it has to diffuse. So it diffuses at a rate proportional to its popularity. However, it competes with all other ideas, which also diffuse at a rate proportional to their popularity – so to write the diffusion rate it will be necessary to divide the popularity of each idea by the "total popularity" in the system. If exactly one agent-idea relationship is added per time period, the total of all popularity is the number of periods, t. Hence k_j, the popularity of idea j born at time t_j, evolves as follows:

$$\dot{k}_j(t) = (1-b)k_j(t)/t.$$

---

[2]I regard the innovation/imitation choice as exogenous, because the forces determining choice can be modelled *independently*, that is, there exists several choice theories compatible with the innovation/diffusion process that I describe.

[3]Throughout the paper, the popularity of an idea is the number of times it is known, that is, the number of agents who have adopted/learned/imitated it.

The factor $(1 - b)$ has been added because I assume that a fraction b of time is spent on innovation, which limits the speed of diffusion. Using the initial condition $k_j(t_j) = 1$ (j is invented by one agent, at some time t_j) this differential equation has solution:

$$k_j(t) = \left(\frac{t}{t_j} \right)^{1-b}. \qquad (4.1)$$

Knowing when ideas are born and their popularity, one can tell, at any point in time, how many of them have a certain popularity. Indeed, the share of ideas known k times, denoted $p(k)$, can be found starting from the cumulative distribution function:

$$Pr(k_j \leq k) = p \left(\left(\frac{t}{t_j} \right)^{1-b} \leq k \right) = 1 - p \left(t_j \leq t k^{-1/(1-b)} \right).$$

Assume that ideas arrived sequentially, in such a way that the t_j's are uniformly distributed i.e. $prob(t_j = Y) = 1/t$ for Y from 1 to t, so $prob(t_j \leq Y) = \sum_1^Y \frac{1}{t} = \frac{Y}{t}$. This leads to $p(k_j \leq k) = 1 - k^{\frac{-1}{1-b}}$. Apply $p(k) = \frac{dp(k_j \leq k)}{dk}$ to retrieve the probability distribution of ideas' popularity,

$$p(k) = \hat{b} k^{-1-\hat{b}},$$

where $\hat{b} = 1/(1 - b)$. It is easy to check that $\int_1^\infty p(k)dk = 1$. This is a power law which steepens with b. The power law exponent is best rewritten $\gamma = 2 + \frac{b}{1-b}$ to show that for $0 < b < 1$ it is greater than 2, and depends positively on the ratio of the share of innovation over the share of diffusion.

The heuristic description above does not account properly for the finiteness of the population, and therefore fails to feature an S-shaped diffusion pattern (see equation 4.1). It does not include the structure of social interactions, and is deterministic. I describe below a more complete mathematical model and its numerical (agent-based) simulation.

4.3.1   The algorithm

Consider a two mode network with n agents and w ideas. Ideas are either known or unknown by any given agent, which is represented by the presence or absence of a link between an agent and an idea. The number of agents is kept fixed, but the number of ideas grows. Time is discrete and indexed by t. Denote by E_t the total number of agent-ideas relationships, i.e. the number of edges of the two-mode network. At the beginning ($t = 1$), there is one idea known by one randomly chosen (r.c.) agent ($w_1 = 1$ and $E_1 = 1$). Then at each time period, as illustrated in figure

4.1, the following algorithm is applied (where random always means uniformly at random):

I/ Pick an agent i at random.

II/ With probability b, the agent i creates a new idea (a new node is added to the set of ideas, and an edge is added to the two-mode network, between i and the new node);

III/ otherwise (i.e. if the r.c. agent does not create a new idea), pick another agent i' at random. Then pick at random an idea j among those ideas known by i' and unknown by i. Then i learns j (an edge is added to the two-mode network, between i and j).[4]

The following section clarifies the setup of the model by deriving key mathematical relationships implied by the algorithm (I-III).

4.3.2 Preliminary results

This section derives two results which will be necessary later and help understanding the setting. First, it is explained that the network density is constant, which is key to obtain a steady-state distribution, and makes clear the importance of assuming a constant b. Second, an expression is derived for the probability that a r.c. idea is known by two r.c. agents. This will be useful when deriving the chances that an idea is unknown by a r.c. agent but known by his friend, which essentially determines the speed of diffusion of individual ideas and hence the popularity distribution.

Consider a matrix Q which has a fixed number of rows (n agents) and a number of columns that depends on time (w_t ideas). The entries Q_{ij} are equal to one if agent i knows idea j, and zero otherwise. This matrix is the incidence matrix of the two-mode network where agent i is linked to idea j if and only if agent i "knows"

---

[4]If both b and n are very small, there are not enough new ideas to satisfy the number of required learning events. This problematic configuration always happens with non negative probability, and to ensure that the model always run, the computer code is as follows: when a r.c. agent i is supposed to learn but his chosen neighbor has nothing new, i creates a new idea. Again, there will always exist a positive probability to find a (directed) pair that cannot perform the exchange. This probability is small in the region of interest, so I do not include this effect in the derivations. In particular, I consider that a knowledge economy is *defined* for $\mu > 0$ (μ is an increasing function of b and n to be defined later. See *infra* and figure 4.3). Moreover, one can correct the main theoretical result equation 4.10 simply by using the "empirical" (from the simulation) values of $b = w_t/E_t$ and μ (equation 4.4)). This condition $\mu > 0$ illustrates that there cannot exist a knowledge economy in which, at a global level, ideas are imitated faster than they are created. This constraint is due to the assumption of inelastic supply of (cognitive) labor. Still, it is possible for individual agents to imitate faster than they innovate, because one newly created idea can be imitated $(n-1)$ times.
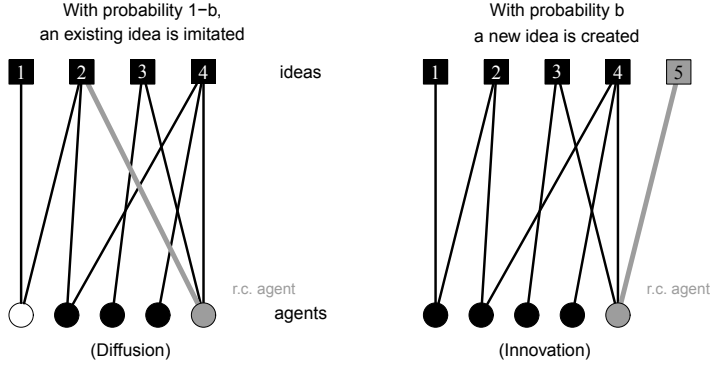
Figure 4.1: Schematic description of the model. At each time step, one and only one of the two events represented above happens. In both cases, a link is added. The main focus here is on the degree distribution of the top nodes (ideas' popularity), $p(k)$. The degree distribution of the bottom nodes is discussed in 4.7.2 but is purposefully uninteresting (agents are homogenous so it is binomial). On the left panel, the r.c. agent (in grey) is learning. In this case, a neighbor has been randomly chosen and turns out to be the leftmost (in white). There are only two ideas unknown by the r.c. agent and known by the r.c. neighbor (1 and 2). The randomly chosen agent chooses uniformly at random an idea of the r.c. neighbor that he doesn't know himself— in the example above he turns out to choose the second idea (a link (in grey) is added between the r.c. agent and this idea). On the right panel, the r.c. agent has created a new idea. The social network between bottom nodes, not depicted here, is assumed to be full throughout the paper except in section 4.5.1.

idea j. Start at $t = 1$ with a column vector filled with a one and $(n-1)$ zeros. At each period, with probability b, a column is added (a new idea is created). Then, with probability 1, one entry of Q is changed from zero to one (if a new column has been added, this modified entry must be in that new column). Since exactly one 1 is added at each period, the total number of ones in Q, which is the total number of edges in the network, is $E_t = t$. The total number of ideas w_t is a random variable equal to W if there has been exactly $W - 1$ successes out of $t - 1$ trials, success happening with probability b. Hence the expected number of ideas is $E(w_t) = 1 - b + bt$. Throughout the paper the concern will be on the long run equilibrium state of the system so I will use $w_t = bt$. Then, it is direct to see

Lemma 4.3.1. *The density of the system, defined as the two-mode network density and denoted D, is stable in the long run:*

Proof.

$$D_t = \frac{E_t}{nw_t} \approx \frac{1}{nb} \quad \square$$

Lemma 4.3.1 shows that if fluctuations due to the stochastic nature of w_t are omitted (which is legitimate in the long run), the density of the two-mode network is constant (independent of system time t), since b and n are fixed parameters. Time independence of the two-mode network density suggests that there may exist a steady state degree distribution. Lemma 4.3.1 shows that an increased rate of innovation b will make the system sparser (since there are more ideas and agents are learning less often), whereas a high rate of learning $(1 - b)$ will make it denser. In this model, growth corresponds to the increment of a column. Diffusion ensures that the density of the system stays stable, by adding positive entries in existing columns.

The key to characterize the self-organized steady-state of the system is to find the number of ideas shared by two r.c. agents, that is, the number of common ideas in a r.c. pair. This is because diffusion takes place between two agents, and is conditioned by what both agents know, since an agent learns only something that his neighbor knows but that he doesn't know himself. Denoting by N_i the set of ideas of agent i and by $|N_i|$ its cardinal, we have

Lemma 4.3.2. *Consider all pairs of agents (i, i') in a system with n agents and w ideas. Then the average (over all pairs) of the number of ideas known by both i and i' is*

$$\langle |N_{i'} \cap N_i| \rangle = \frac{\sum_{i' < i} |N_{i'} \cap N_i|}{\# of pairs} = \frac{\sum_{k=1}^{n} \binom{k}{2} P(k)}{\binom{n}{2}} = \frac{w(\langle k^2 \rangle - \langle k \rangle)}{n(n-1)} \qquad (4.2)$$

Proof. Observe that the sum over all pairs of $|N_{i'} \cap N_i|$ is simply the total number of pairwise "overlaps" in the system, i.e. the total number of times that the triplet "two agents linked to an idea" can be found in the network. Since each idea known k_j times produces $\binom{k_j}{2}$ overlaps between pairs, and denoting $P(k)$ the number of ideas with degree k, the sum can be obtained. Using $P(k) = wp(k)$ and denoting $\langle k^r \rangle = \sum_{j=1}^{w} k_j^r = \sum_{k=1}^{n} k^r p(k)$ gives the simplified form. $\square$

Note that lemma 4.3.2 holds in quite general conditions but gives only the average value of pairwise overlap, not its distribution across different pairs. The average will be very informative because the distribution turns out to be tightly peaked around its mean, since I have excluded all structural sources of agents' heterogeneity. In practice, lemma 4.3.2 will often be used after substituting $\langle k \rangle = E_t/w_t = 1/b$.

The main objective is to derive $p_t(k)$, the probability that a r.c. idea in t is known k times (i.e. has degree k). Under what conditions will idea j be learned at time t? First, the r.c. agent i must be learning, which happens with probability

$(1-b)$. Second, the r.c. pair (i, i') must be such that $j \in N_{i'} \backslash N_i$, that is, j belongs to the set of ideas which are both known by i' and not known by i. Third, idea j must be the one chosen among all other ideas $j' : j' \in N_{i'} \backslash N_i$. At each period, conditional on the event "learning" being realized, exactly one idea must be chosen. The attachment kernel gives the probability that a particular one be chosen, that is

$$A_t(k_j) := Pr(k_j(t+1) = k_j(t) + 1) \; ; \; \sum_{j=1}^{w_t} A_t(k_j) = 1 - b$$

$$A_t(k_j) = (1-b)\Big\langle \frac{Pr(j \in N_{i'} \backslash N_i)}{|N_{i'} \backslash N_i|} \Big\rangle,$$

where the angle brackets denote average over all possible (ordered) pairs. Since agents and pairs of agents are *ex-ante* homogenous, and since there is no source of strong *ex-post* heterogeneity, I simply assume that all pairs have the same value of $Pr(j \in N_{i'} \backslash N_i)$ and $|N_{i'} \backslash N_i|$. Hence,

$$A_t(k_j) = (1-b)\frac{Pr(j \in N_{i'} \backslash N_i)}{|N_{i'} \backslash N_i|}. \tag{4.3}$$

The probability that $j \in N_{i'} \backslash N_i$ can be computed as follows. We want to know the number of choices for finding an ordered pair (i, i') such that $j \in N_{i'} \backslash N_i$. There are k_j choices for i' such that $j \in N_{i'}$ and $(n - k_j)$ choices for i such that $j \notin N_i$. Hence there are $k_j(n - k_j)$ choices for an ordered pair such that $j \in N_{i'} \backslash N_i$. Since there are $n(n-1)$ ordered pairs, a r.c. pair will exhibit $j \in N_{i'} \backslash N_i$ with probability $\frac{k_j(n-k_j)}{n(n-1)}$. The denominator $|N_{i'} \backslash N_i|$ can be computed as $|N_{i'} \backslash N_i| = |N_{i'}| - |N_{i'} \cap N_i|$. Using again the assumption that this is the same for all pairs, $|N_{i'} \backslash N_i| = \langle |N_{i'}| \rangle - \langle |N_{i'} \cap N_i| \rangle = \frac{\mu t}{n-1}$, where equation 4.2 was used and μ is defined as

$$\mu(t) := 1 - \frac{w_t}{E_t} \frac{\langle k^2 \rangle}{n} = 1 - \frac{\langle (k/n)^2 \rangle}{D_t}. \tag{4.4}$$

The last step in equation 4.4 uses $w_t = bt$, $E_t = t$ and lemma 4.3.1. I omit the time subscript in $\langle k^2 \rangle = \sum_{j=1}^{w_t} [k_j(t)]^2$. The attachment kernel equation 4.3 can now be written

$$A_t(k_j) = \frac{k_j(n - k_j)}{\hat{b}\mu n t}. \tag{4.5}$$

The condition $\sum_{j=1}^{w_t} A_t(k_j) = 1 - b$ is the same equation as the definition of μ (equation 4.4). μ ensures that the attachment kernel is correctly normalized, that is, if the event of period t is imitation, the chances that a particular idea diffuses

are such that exactly one will diffuse. In this sense, μ characterizes the degree of competition among ideas. The higher μ, the lower the chances that each particular idea diffuses. μ indicates how many ideas are available for diffusion, in a precise sense. Since the chances of "meeting" an unknown idea j is the number of times that j is known by somebody else (or by a friend, if the friendship network is sparse), at this level each idea competes with all ideas *unknown* by a r.c. agent (not with all other ideas in the system). Algebraically, μ as defined in equation 4.4 admits the following combinatorial interpretation

Proposition 4.3.1. *μ is the average of the individual quantities μ_i, where μ_i is the fraction of edges that are pointing to ideas unknown by agent i.*

$$\mu = \frac{1}{n}\sum_{i=1}^{n}\mu_i \ ; \ \ \mu_i = \frac{\sum_{j\notin N_i}k_j}{\sum_{j=1}^{w}k_j} \tag{4.6}$$

Proof. The denominator of μ_i is simply the total number of edges, E_t. The numerator of μ_i can be rewritten $\sum_{j\notin N_i}k_j = \sum_{j=1}^{w}k_j(1-Q_{ij})$ where Q_{ij} are the entries of the incidence matrix, equal to one if i knows j and zero otherwise. Hence,

$$\mu = \frac{1}{nE_t}\sum_{i=1}^{n}\sum_{j=1}^{w}[k_j(1-Q_{ij})].$$

Transposing the two sums and decomposing the sum over i, this becomes

$$\mu = \frac{1}{nE_t}\sum_{j=1}^{w}[\sum_{i=1}^{n}k_j - \sum_{i=1}^{n}k_jQ_{ij}].$$

It is easy to see that by definition $\sum_{i=1}^{n}k_j = nk_j$ and $\sum_{i=1}^{n}k_jQ_{ij} = k_j^2$. Therefore,

$$\mu = \frac{1}{nE_t}\sum_{j=1}^{w}\left[nk_j - k_j^2\right] = 1 - \frac{w_t}{E_t}\frac{\langle k^2\rangle}{n} \quad \square$$

The factor $\mu(t)$ is defined at all periods of time and helps characterizing the dynamics of the system. However, it depends itself on the dynamics of the system. How is this feedback loop solved? Does the system stabilize? Since the distribution $p(k)$ depends on the attachment kernel, and the attachment kernel depends on μ which depends on the second-order moment of the distribution, equation 4.4 is a fixed point equation, i.e. $\mu = f(\mu, b, n, t)$. If the popularity distribution is stable, its second order moment is stable and so is μ.

I show below that assuming that μ is constant and that a steady-state exists, the steady-state is unique. This gives a steady-state value of $\langle k^2\rangle$, which can be

inserted into equation 4.4 to obtain a steady-state fixed point equation for μ.

4.4   Results

4.4.1   Distribution of ideas' popularity

In view of the attachment kernel (4.5), the flows in and out of the kth bin of the histogram can be written explicitly, following the method of Simon. Recall that $P_t(k)$ is the total number of ideas with degree k at time t. Then,

$$P_{t+1}(k) - P_t(k) = P_t(k-1)A_t(k-1) - P_t(k)A_t(k).$$

Using $P_t(k) = btp_t(k)$ and $A_t(k)$ from equation 4.5

$$t\Big(p_{t+1}(k) - p_t(k)\Big) + p_{t+1}(k) = p_t(k-1)\frac{(k-1)(n-(k-1))}{\hat{b}\mu n} - p_t(k)\frac{k(n-k)}{\hat{b}\mu n}.$$

Assuming a steady state in the sense that $p_{t+1}(k) = p_t(k) = p(k)$ gives the recurrence

$$p(k)(k(n-k) + \hat{b}\mu n) = p(k-1)(k-1)(n-(k-1)). \tag{4.7}$$

Equation 4.7 can be iterated to give

$$p(k) = p(1)\prod_{i=1}^{k-1} \frac{i(n-i)}{\hat{b}n\mu + (i+1)(n-(i+1))}. \tag{4.8}$$

Making use of the quadratic formula, the denominator can be rewritten $(-1)(i - u_1)(i-u_2)$ where $\{u_1, u_2\} = \frac{1}{2}\left(n-2 \pm \sqrt{n(n+4\hat{b}\mu)}\right)$. Now consider the definition of the Pochhammer symbol:

$$(x)_y = x(x+1)(x+2)\dots(x+y-1) = \frac{\Gamma(x+y)}{\Gamma(x)}. \tag{4.9}$$

Expanding the product in (4.8) and using (4.9) on each of the terms gives

$$p(k) = p(1)\frac{(1)_{k-1}(n-(k-1))_{k-1}}{(-1)^{k-1}(1-u_1)_{k-1}(1-u_2)_{k-1}}.$$

From Slater's (1966) formula I.5 p. 239, $(n-(k-1))_{k-1} = (-1)^{k-1}(1-n)_{k-1}$. Therefore,

Proposition 4.4.1. *The steady-state distribution of ideas' popularity is given by*

$$p(k) = p(1)\frac{(1)_{k-1}(1-n)_{k-1}}{(r_1)_{k-1}(r_2)_{k-1}}, \tag{4.10}$$

where

$$\{r_1, r_2\} = \frac{4 - n \pm \sqrt{n(n + 4\hat{b}\mu)}}{2}$$

and

$$p(1) = \left(1 + \frac{n-1}{n\hat{b}\mu}\right)^{-1}. \tag{4.11}$$
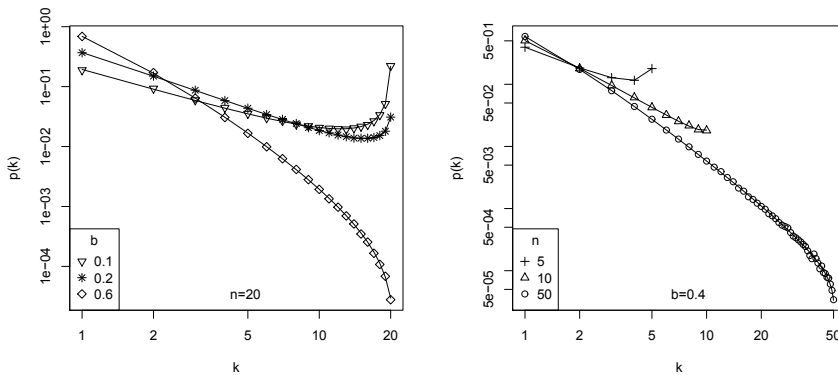
Figure 4.2: Distribution of ideas' popularity. For each of six configurations of parameters, the model is run once for 3×10^6 periods (only 10^6 when $n = 5, n = 10$). In the left panel, $n = 20$ and the effect of b is studied. In the right panel, $b = 0.4$ and the effect of n is studied. The plain lines are the theoretical results, computed using equation 4.10 and values of μ computed using the fixed point equation (4.14). These six points of the parameter space are marked in figure 4.3. When a point in figure 4.3 is in the lower left half of the (μ, b) plane, the corresponding curve in the figures above exhibit an upward curvature, otherwise it exhibits a downward curvature.

The term $p(1)$ is found by setting up the appropriate master equation, in which there are no inflows from the 0th bin but there is a probability of innovation: $P_{t+1}(1) - P_t(1) = b - P_t(1)A_t(1)$. Assuming a steady-state and solving for $p(1)$ gives (4.11).

The probability mass function (4.10) is plotted against simulations in figure 4.2. In some region of the parameter space, it has an upward curvature in the tail.[5] This

_____

[5] A similar phenomenon was found by Peruani et al. (2007) on the degree distribution of the *fixed* set of nodes, in a growing two-mode network with mixed (random and preferential) attachment, and a high value of the parameter tuning the relative amount of preferential versus random attachment. The model of Evans & Plato (2008), which is a fixed two-mode network with rewiring,

curvature exists when the function admits a minimum at some $k = k^* < n$. Using (4.10), the point at which $p(k^*) = p(k^* - 1)$ is given by $k^* = \frac{1}{2}(1 + n(1 + \hat{b}\mu))$ and the point at which $p(k^* + 1) = p(k^*)$ is given by $k^* = \frac{1}{2}(-1 + n(1 + \hat{b}\mu))$ so that we may take $k^* = \frac{1}{2}n(1 + \hat{b}\mu)$. The condition $k^* < n$ is then the same as $\mu < 1 - b$. The region of the parameter space for which this condition holds, such that an upward curvature exists, is the lower left half of figure 4.3 (see section 4.4.2). The latter corresponds to relatively low values of b and n (but conditional on b and n being large enough to have $\mu > 0$; see footnote 4).

To obtain further insights onto the nature of the distribution (4.10), consider verifying that the terms sum up to one. These terms are hypergeometric, so the sum is of the form

$$\sum_{k=1}^{n} p(k) = p(1) \sum_{k=1}^{n} \frac{(1)_{k-1}(1-n)_{k-1}}{(r_1)_{k-1}(r_2)_{k-1}} = p(1) \, _3F_2[\{1, 1, 1 - n\}, \{r_1, r_2\}, 1].$$

The five parameters of this generalized hypergeometric function ($_3F_2[\ldots]$) satisfy an important constraint. This $_3F_2$ is $1-$balanced, that is, its parametric excess is equal to one:

$$(r_1 + r_2) - (1 + 1 + (1 - n)) = 1.$$

It means that this $_3F_2$ is Saalschützian. Hence, the Pfaff-Saalschütz summation theorem can be applied to check that (4.10) and (4.11) define a properly normalized probability mass function

$$_3F_2[\{1, 1, 1 - n\}, \{r_1, r_2\}, 1] = \frac{(r_1 - 1)_{n-1}(r_1 - 1)_{n-1}}{(r_1)_{n-1}(r_1 - 2)_{n-1}} = \frac{n(1 + \hat{b}\mu) - 1}{\hat{b}\mu n} = 1/p(1).$$

Note that many other distributions are, in this sense, Pfaff-Saalschützian. More generally, the steady state distribution (4.10) is a generalized hypergeometric probability distribution (GHPD). It is named so because its generating function is a ratio of generalized hypergeometric functions (Johnson et al. 2005). In the case of (4.10), the generating function takes the following particular form.[6]

$$G(z) = \sum_{k=1}^{n} p(k) z^k = \frac{_3F_2[\{1, 1, 1 - n\}, \{r_1, r_2\}, z]}{_3F_2[\{1, 1, 1 - n\}, \{r_1, r_2\}, 1]}. \tag{4.12}$$

This class is interesting because there exists a deep connection between Pfaff-Saalschütz and Gauss hypergeometric theorems, and Gauss hypergeometric function

---

can also produce a U-shaped distribution, when the relative amount of preferential v.s. random attachment is high.

[6]More general cases involving 5-parameters generalized hypergeometric functions are given in Johnson et al. (2005) and Gutiérrez Jáimez & Rodríguez Avi (1997).

is the generating function of, *inter alia*, the Poisson, binomial, negative binomial, hypergeometric, and Waring distribution. In fact each theorem can be obtained starting from the other (Slater 1966, p. 48-49). The convergence of Pfaff-Saalschütz to Gauss theorem, applied to the finite population distribution (4.10), shows that

Proposition 4.4.2. *For $n \to \infty$, the distribution of ideas popularity is the Yule-Simon distribution*

$$p(k) = \hat{b}B(k, \hat{b} + 1), \tag{4.13}$$

where $B()$ is the beta function.[7] *The condition $\sum_{k=1}^{\infty} p(k) = 1$ can be verified using Gauss hypergeometric theorem.*

Proof. Consider the limit of each term of (4.10). Assuming $\lim_{n\to\infty} \mu = 1$, as will be justified in section 4.4.2, $\lim_{n\to\infty} p(1) = \frac{\hat{b}}{\hat{b}+1}$. Also, $\lim_{n\to\infty} r_1 = 2 + \hat{b}$. Furthermore, $2 - n - \hat{b}\mu < r_2 < 2 - n$ and (Slater 1966, p.49) $\lim_{n\to\infty} \frac{(1-n)_{k-1}}{(2-n-\hat{b}\mu)_{k-1}} = \lim_{n\to\infty} \frac{(1-n)_{k-1}}{(2-n)_{k-1}} = 1$ so $\lim_{n\to\infty} \frac{(1-n)_{k-1}}{(r_2)_{k-1}} = 1$. Combining all three limits, $\lim_{n\to\infty} p(k) = \frac{\hat{b}}{\hat{b}+1} \frac{(1)_{k-1}}{(2+\hat{b})_{k-1}} = \frac{\hat{b}}{\hat{b}+1}(1 + \hat{b})B(k, 1 + \hat{b})$ which simplifies to (4.13) $\qquad\square$

It should be emphasized that this result is not a steady-state result. The steady-state result for all values of n is given in proposition 4.4.1. However, when n is very large and t is not large enough for finite-size effects to be observed (no idea has had enough time to diffuse to the whole population), the Yule-Simon gives a good approximation of the observed distribution.

A last remark on the distribution (4.10) is its relation to the beta distribution. In the mean field-deterministic-continuous approximation of the stochastic process, the variable k/n follows a distribution proportional to $(k/n)^{-1-\hat{b}\mu}(1-k/n)^{-1+\hat{b}\mu}$ (see 4.7.1). However, the support is on $[1/n, 1]$ instead of $[0, 1]$ for the classical beta distribution, and the restriction on the parameters in the beta (both parameters must be positive) does not hold. The mean-field deterministic approximation is also useful to see that the (expected) diffusion is S-shaped (equation 4.17).

The distribution (4.10) is not fully closed form, in the sense that the term μ appears in it, while also depending on it. I now turn to determining the steady-state value of μ.

---

[7]The beta function is defined in terms of the Gamma function: $B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$. The Gamma function generalizes the factorial function for non integer values, such that when x is an integer $\Gamma(x + 1) = x\Gamma(x) = x!$, but x can also take non-integer values. It relates to the Pochhammer symbol through equation 4.9.

4.4.2   Properties of the partition factor

Practically, to compute the predicted steady-state distribution, the value $\mu(b, n)$ is needed. This value can be recorded from the simulations, using either (4.4) or (4.6), which are equal by proposition 4.3.1. However, it is also possible to compute, prior to the simulations, the tables of μ at its steady-state (so that (4.10) is genuinely closed-form), for all values of b and n. The steady-state value of μ attained by the stochastic system turns out to be unique, even though the self-consistency fixed point equation studied below admits a second fixed point in the interval of interest. The second fixed point is $\mu = 1 - b$ for all values of n and can be formally proven.[8] As already mentioned, this fixed point separates the two regions of the parameter space for which there exists or not an upward curvature in the steady state distribution (4.10).

In the general case the objective is to solve equation 4.4 for μ with $\langle k^2 \rangle$ taken at its steady state value. The steady-state value of $\langle k^2 \rangle$ and other moments are readily determined

Proposition 4.4.3. *The moments of the popularity distribution are*

$$\langle k^r \rangle = p_{1\ r+1}F_r[\{2, 2, 2, \ldots, 1 - n\}, \{1, 1, \ldots, r_1, r_2\}, \{1\}].$$

Proof. Each successive term is found by multiplying by $k = \frac{(2)_{k-1}}{(1)_{k-1}}$.   □

Inserting the steady-state value of $\langle k^2 \rangle$ and w_t in equation 4.4 gives the fixed point equation

$$\mu = 1 - \frac{b}{n}\ p_{1\ 3}F_2[\{2, 2, 1 - n\}, \{r_1, 4 - n - r_1\}, 1]. \qquad (4.14)$$

This equation is solved numerically in the region of interest ($b \in\ ]0, 1[$). I computed values of f_μ (the RHS of the equation) for 99 values of b and a few values of n, and then obtained the fixed points by studying at which points $\mu - f_\mu$ changes sign. The results are reproduced in figure 4.3 where one can see, abstracting from the $\mu = 1 - b$ line, that μ is monotonically increasing and concave in b and n. When the population is large, or when innovation is high, an agent i knows only a small proportion of all ideas, and hence the popularity of the ideas unknown by i is high

---

[8] Upon substituting $\mu = 1 - b$, which cancels b, one obtains the surprising one-parameter generalized hypergeometric function identity

$$_3F_2[\{2, 2, 1 - n\}, \{\tfrac{1}{2}\left(4 - n + \sqrt{n(n + 4)}\right), \tfrac{1}{2}\left(4 - n - \sqrt{n(n + 4)}\right)\}, 1] = 2n - 1.$$

It can be proven using the computer implementation of Gosper's (1978) algorithm by Paule & Schorn (1995). On this topic, see Petkovšek et al. (1996).
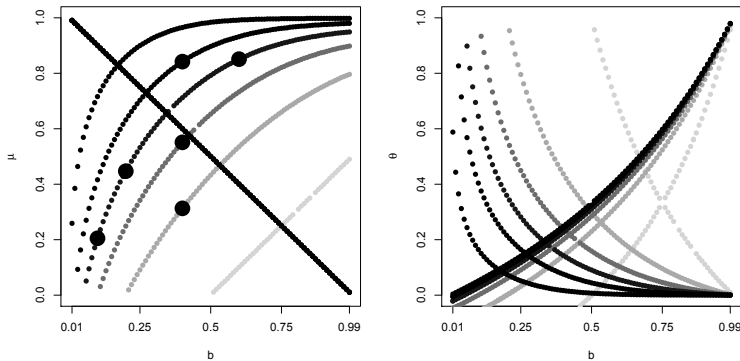
Figure 4.3: Left panel: Numerically computed fixed points of equation 4.4 at the steady-state (i.e. solutions of equation 4.14). From the clearest to darkest points, $n = 2, 5, 10, 20, 50, 500$. The six large black dots correspond to the six points of the parameter space used in figure 4.2. Their position with respect to the line $\mu = 1 - b$ (above or below) determines the shape of the curvature in figure 4.2 (downward or upward). Note that darker dots overlap brighter ones along the line $\mu = 1 - b$. Right panel: The decreasing curves represent the average overlap θ computed using equation 4.15 (the increasing curves correspond to the fixed point $\mu = 1 - b$).

as compared to the total popularity of all ideas. By proposition 4.3.1, this implies a higher μ.

For small values of n, μ can be found explicitly but at considerable computational cost. It involves solving polynomials of the order of n. As it turns out, this polynomial always has a root in $\mu = 1 - b$. When $n = 2$, appendix 4.7.3 shows that the other root is

$$\mu(n = 2) = -\frac{1}{2} + b.$$

Finally, note that the values of μ recorded directly from the simulations (unreported) are in good agreement with the numerical solution of equation 4.14. However, for values of μ close to 0, a significant departure can be observed, especially for low n. This is due to the fact that in these cases, many innovation events occur because of learning events failing (the chosen friend does not have any original ideas to offer, see footnote 4). So, in these cases, the recorded value of μ is not in very good agreement with the input value of b, chosen in advance of the simulation as a parameter; however, it is in excellent agreement with the effective value of b computed as $b = w_t/t$.

It should be emphasized that the convergence of $\mu(t)$ to a fixed point indicates the self-organization of the system. Self-organization results from the feedback loop between structure and dynamics, which comes from the fact that what is learned depends on what is known/unknown. Because μ is a structural quantity capturing the organization of who knows what (proposition 4.3.1), and because it determines who learns what (equation 4.5), it is a fundamental quantity. Its convergence to a

fixed point reveals that at this level of aggregation, the system self-organizes into a stable state. To understand this more intuitively, the next section shows that the average pairwise overlap is related to μ.

4.4.3   Average overlap

Consider the average overlap between two given agents, defined as the Jaccard index of their knowledge portfolios

$$\theta_{ii'} = \frac{|N_i \cap N_{i'}|}{|N_i \cup N_{i'}|} = \frac{|N_i \cap N_{i'}|}{|N_i| + |N_{i'}| - |N_i \cap N_{i'}|}.$$

The average over all pairs of agents is

$$\theta = \langle \theta_{ii'} \rangle \approx \frac{\langle |N_i \cap N_{i'}| \rangle}{\langle |N_i| + |N_{i'}| - |N_i \cap N_{i'}| \rangle} = \frac{\langle |N_i \cap N_{i'}| \rangle}{2 \langle |N_i| \rangle - \langle |N_i \cap N_{i'}| \rangle}.$$

The first relationship is not exact because the expectation of a ratio is, in general, different from the ratio of expectations. However, pairs are very similar in terms of the sizes of their intersections and unions, so that the distribution of these sizes are very tightly peaked, making the approximation fairly good. Now we can use $\langle |N_i| \rangle = t/n$, lemma 4.3.2 and equation 4.4, to get

Proposition 4.4.4. *The average overlap between agents is well approximated by*

$$\theta = \frac{1 - \mu - 1/n}{1 + \mu - 1/n}. \tag{4.15}$$

Since μ is monotonically increasing in b, the average overlap θ decreases with innovation and increases with learning. Intuitively, an agent who learns ideas of others gets closer to them, and an agent who invents his own ideas increases his distinctiveness. It can also be seen in figure 4.3 (right panel) that θ is also decreasing in n, because it is harder to maintain a high overlap with everybody when there are many agents.

Since there is a one-to-one mapping between b and μ, equation 4.15 implies a one-to-one mapping between b and θ. Hence, for a given number of agents, the rate of innovation determines the average overlap between two agents' portfolio. If the model is reversed in the sense that agents choose to imitate or innovate so as to have a certain θ^*, then, given n, the effective $b = w_t/E_t$ is uniquely determined. In other words, while assuming a fixed innovation-imitation trade-off produces a certain (self-)organization, assuming a certain self-organization would determine the innovation-imitation trade-off. This of course would depend on the particular

assumptions made, and would require further work to be rigorously analyzed.[9] However, this remark helps to emphasize that in this model, if there exists an optimal overlap θ^* there also exists an optimal rate of innovation b^*.

4.5 A few generalizations

4.5.1 Social network

The derivation of the distribution (4.10) was made by assuming a complete social network. Consider an opposite case.

Proposition 4.5.1. *If the social network is a circle in which agents have one friend on each side, the distribution is geometric (with a slight modification for $p(n)$):*

$$p(k) = b(1-b)^{k-1} \qquad for \ k \in [1, n-1],$$
$$p(n) = (1-b)^{n-1}.$$

Proof. See Appendix 4.7.4.

For other types of social networks, simulation results are reported in figure 4.4. If we stay with a one dimensional circular lattice, as in proposition 4.5.1, but with a larger number of neighbors on each side, this creates the possibility for an idea to be known by two neighbors of an agent, and the derivation above becomes inexact. However, this configuration would not happen very often, so that for circle networks with small degree, the distribution stays geometric (panel f on figure 4.4). However, when the number of neighbors increases to a maximum, the network becomes complete, so that the degree distribution converges to the one obtained under the complete network assumption (see panel d). Note that the important criterium to determine the shape of the popularity distribution is not the average degree of an agent, because the competition among ideas cancels out this effect. For instance, panels a and b show that even sparse Erdős-Renyi networks give results roughly similar to complete networks.[10] The decisive criteria is the dependence

---

[9]Using simulations, I checked the following. At each period, a directed pair of agents is chosen at random. Agent i computes $\theta_{ii'}$ and learns from i' if $\theta_{ii'}$ is less than some predetermined θ^*, otherwise he innovates. As expected, in the simulations explored, pairs of agents converge to $\theta_{ii'} \approx \theta^*$ and the system itself exhibits $w_t/t = b_{effective} \approx b_{theory}$ where b_{theory} is computed by inverting the relationship between θ and b through μ.

[10]There was a probability that these random networks would not be fully connected. I forced connectedness by repeating the algorithm of graph creation until one fully connected was created – so strictly speaking, they are not Erdős-Rényi but random graphs from a (slightly) restricted ensemble.
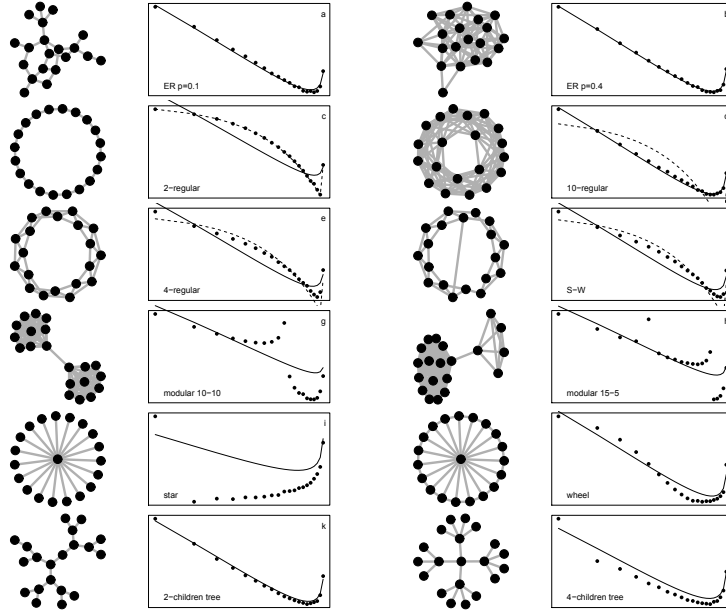
Figure 4.4: Simulations for 10^6 periods, using different social networks, $n = 20$ and $b = 0.2$. The plain line is the theoretical result for a complete social network, as in figure 4.2. The dashed line is the theoretical result for one-dimensional 2-regular circular lattice (proposition 4.5.1). The first line of the panel shows two (connected) Erdős-Rényi ($p = 0.1, 0.4$). The second line shows two q-regular circular one dimensional lattice ($q = 2, 10$). The third line shows two 4-regular one dimensional circular lattice, with one rewired edge for the right one. The fourth line shows two modular networks, constructed by linking two complete subgraphs. The fifth line shows a star network, and a combination of a star and a one dimensional 2-regular circular lattice. The last line shows two regular trees, one with two children, and one with 4 children (some nodes have less children, due to the requirement that $n = 20$).

or independence of the attachment kernel on k_j, that is, the fact that the rate of diffusion of an idea depends or not on its popularity. While relating arbitrary social network structure to the popularity distribution by analytical methods is out of the scope of this chapter, it can be argued heuristically that social networks with a relatively high number of short cycles will tend to produce an attachment kernel which is not preferential, whereas networks which are closer to trees, such as Barabási-Albert or Erdős-Rényi networks, will tend to produce attachment kernel which are preferential. This comes from the fact that when there are short cycles in the social network, ideas do not diffuse mostly to agents whose neighbors are ignorant of that idea.

4.5.2 Differentiated productivity

This section relaxes the unrealistic assumption that conditional on investing one unit of time, agents get as many ideas by learning as by innovating. Instead of learning or creating one single idea, agents now have a fixed productivity. When they innovate, they create λ_P ideas, and when they learn, they learn λ_L ideas (sampling a new neighbor with replacement every time).[11] The attachment kernel is now given by

$$A_t(k) = (1-b)\lambda_L \frac{P(j \in N_{i'}\backslash N_i)}{\sum_j P(j \in N_{i'}\backslash N_i)},$$

where $P(j \in N_{i'}\backslash N_i) = \frac{k(n-k)}{n(n-1)}$ does not change. The productivity of learning does not change the nature of the diffusion process, but simply its speed. The productivity of innovation now determines the total number of ideas, $w_t = b\lambda_P t$, and the total number of edges $E_t = t(b\lambda_P + (1-b)\lambda_L)$. It still holds that $\sum_{j=1}^{w} P(j \in N_{i'}\backslash N_i) = \frac{nE_t - w_t\langle k^2\rangle}{n(n-1)}$ so that

$$A_t(k) = (1-b)\lambda_L \frac{k(n-k)}{nE_t - w_t\langle k^2\rangle} = \frac{k(n-k)}{(\zeta+1)\mu nt},$$

where μ is still defined by equation 4.4, and the combinatorial interpretation (proposition 4.3.1) still holds. The parameter ζ is defined as $\zeta = \frac{b\lambda_P}{(1-b)\lambda_L}$. Note that if we set $\lambda_P = \lambda_L = 1$, we find $\zeta + 1 = \hat{b}$ as it must to recover the attachment kernel (4.5). The procedure to find the steady-state distribution (section 4.4.1) can be followed here as well. The resulting degree distribution simply now balances the rate of innovation $b\lambda_P$ with the rate of learning $(1-b)\lambda_L$ (instead of only b with $1-b$). In the limit of an infinite population, the exponent of the Yule-Simon was $2 + \frac{b}{1-b}$, and with productivity parameters it can be shown that it is $2 + \zeta$. This highlights that the original and productivity-augmented models can really be thought of as one parameter (ζ) models.

4.6 Conclusion

The importance of innovation and knowledge diffusion in economic systems is widely recognized. Likewise, the literature has emphasized the role of interactions and self-reinforcing dynamics in shaping the structure and dynamics of economies. This context calls for a fundamental understanding of the self-organization of knowledge economies: considering interacting agents who innovate new-to-the-world ideas and

---

[11]λ_L must be a small number to ensure that there are enough ideas to be learned. See footnote 4.

imitate existing ideas, what can we say about the likely long term structure of who knows what?

In this paper, I have characterized a parsimonious model of knowledge diffusion and growth. In the model, learning ideas of friends implies self-reinforcing but bounded diffusion, leading to an S-shaped or logistic diffusion curve. Together with a continuous arrival of new ideas, a stable organization emerges in terms of the distribution of ideas' popularity. In general, because who knows what determines who learns what, there is a feedback loop between the structure and dynamics of the system. In the model studied, this feedback loop leads to a stable state which is thus "self-organized". Moreover, it is shown that under most circumstances, the distribution of ideas' popularity is close to a power law: most ideas are known by only a few agents, who just discovered them, and only a few ideas have diffused completely. A higher rate of diffusion, relative to the rate of innovation, implies a fatter right tail of the popularity distribution (more diffusion implies more very well known ideas). Moreover, the structure of the social network on which ideas' diffusion takes place matters in a non trivial way. Heuristically, networks with short cycles, as opposed to tree-like networks, prevent ideas from diffusing as fast as they could, leading to less skewed distribution of ideas's popularity.

The main consequence of these results is that in a society which facilitates relatively more diffusion than innovation (which implies a high λ_L, and a low b if the choice of innovation/ imitation depends on the relative returns to each activity), we should expect the distribution of ideas' popularity to be very skewed and the average overlap to be very high. On the other hand, in a society which favors the emergence of genuinely new ideas, we should expect the distribution of ideas' popularity to fall faster, and the average overlap to be lower.

Finally, the model or some technical aspects of this paper may be of use in other applications. Since the model presented is a non-linear (logistic) extension of the widely used model of Simon's (1955), and given the large number of phenomena across which power laws and logistic growth are observed, it might be of interest beyond the literature on knowledge systems, and it contributes to the ongoing research agenda on the evolution of networks.

4.7 Appendix

4.7.1 Distribution of ideas' popularity: mean-field continuous deterministic approximation

Consider that each idea j diffuses deterministically and assume that time is continuous. Using (4.5),

$$\frac{dk_j(t)}{dt} = \frac{k_j(t)(n - k_j(t))}{\hat{b}\mu n t}.$$
(4.16)

This is a first-order ordinary differential equation. It looks similar to Verhulst's equation of population growth, except that it has non constant coefficients since t appears on the RHS. It is non-linear, but it is a Bernoulli equation so it can be linearized and integrated. We could also note that it is an exact differential equation and apply relevant techniques. The simplest is probably to separate variables to obtain

$$\hat{b}\mu n \int \frac{1}{k(n-k)} dk = \int \frac{1}{t} dt$$

$$\hat{b}\mu \left[\log\left(\frac{k}{k-n} \right) + C_1 \right] = \log(t) + C_2$$

$$k_j(t) = n/(1 - Ct^{-1/\hat{b}\mu}),$$

where C is an arbitrary constant. Using the initial condition $k_j(t_j) = 1$, it follows that $C = (n-1)/(t_j^{-1/\hat{b}\mu})$, and therefore the solution of (4.16) is

$$k_j(t) = n \left[1 + (n-1) \left(\frac{t_j}{t} \right)^{\frac{1-b}{\mu}} \right]^{-1}.$$
(4.17)

Note that equation 4.17 is a logistic curve, that is, diffusion is S-shaped (from equation 4.16, $\frac{d^2 k_j(t)}{dk^2}$ changes sign at $k_j = n/2$). The continuous distribution is computed thus (using equation 4.17):

$$p(k_j \leq k) = p \left(n \left[1 + (n-1) \left(\frac{t_j}{t} \right)^{\frac{1-b}{\mu}} \right]^{-1} \leq k \right)$$

$$= 1 - p \left(t_j \leq \left(\frac{k_j - k_j n}{k_j - n} \right)^{\hat{b}\mu} t \right).$$

Since the t_j's are uniformly distributed[12] their probability mass function is $Pr(t_j = Y) = 1/t$ for Y from 1 to t, so $Pr(t_j \leq Y) = \sum_1^Y \frac{1}{t} = \frac{Y}{t}$. This leads to

$$p(k_j \leq k) = 1 - \left(\frac{k_j - k_j n}{k_j - n} \right)^{\hat{b}\mu}.$$

Applying $p(k) = \frac{dp(k_j \leq k)}{dk}$ gives

$$p(k) = \hat{b}\mu n(n-1)^{-\hat{b}\mu} (n-k)^{-1+\hat{b}\mu} k^{-1-\hat{b}\mu}. \tag{4.18}$$

One can check that this is a proper distribution function, $\int_1^n p(k)dk = 1$. This distribution has the shape of a particular beta distribution. Making the change of variable $x = k/n$, we have

$$p(x) \propto (1-x)^{-1+\hat{b}\mu} x^{-1-\hat{b}\mu},$$

which is almost the definition of a beta distribution beta(α, β) with $\alpha = -\hat{b}\mu$ and $\beta = \hat{b}\mu$. However, negative parameters are not allowed in the definition of the beta distribution. Moreover, the factor of proportionality is different from that of the beta distribution because the support is different. This distribution has to have a strictly positive support, because the integral diverges at 0.

4.7.2   Distribution of agents' number of ideas known

Below it is shown that the the number of ideas known by a r.c. agent has a binomial distribution. To "know" k_a ideas at time t, a r.c. agent needs to have been chosen exactly k_a times, and not chosen exactly $(t - k_a)$ times. Thus it follows that the distribution of agents' number of ideas known is the binomial distribution

$$p_t(k_a) = \binom{t}{k_a} \left(\frac{1}{n} \right)^{k_a} \left(1 - \frac{1}{n} \right)^{t-k_a}.$$

4.7.3   Exact solution of the fixed point equation for $n = 2$

Written explicitly the fixed point equation 4.14 becomes

$$\mu = 1 - \frac{b}{n}p(1) \sum_{k=1}^n k^2 \frac{(1)_{k-1}(1-n)_{k-1}}{(r_1)_{k-1}(4-n-r_1)_{k-1}}. \tag{4.19}$$

---

[12]Contrary to one-mode scale free network models, this is not exactly true, since there is not one new idea per period, but only one at each period with probability b. The uniform distribution is, nevertheless, an appropriate approximation since the t_js of many independent realizations of the stochastic process are uniformly distributed over $[1, t]$.

For $n = 2$, the series has only two terms and $r_1 r_2 = -2\hat{b}\mu$. The series itself is thus $1 + 4\frac{1-2}{r_1 r_2} = \frac{\hat{b}\mu+2}{\hat{b}\mu}$. The term $p(1)$ is $(p_1|n = 2) = \frac{2\hat{b}\mu}{2\hat{b}\mu+1}$. Substituting into (4.19),

$$1 - \mu = \frac{b(2\hat{b}\mu)(\hat{b}\mu - 2)}{2(2\hat{b}\mu + 1)a} = \frac{b(\hat{b}\mu + 2)}{2\hat{b}\mu + 1},$$

that is

$$-2\hat{b}\mu^2 + \hat{b}\mu + 1 - 2b = 0$$

$$\mu = \{1 - b, -1/2 + b\}. \tag{4.20}$$

This result can also be derived using the inclusion/exclusion formula or by writing the dynamic process for μ_i.

4.7.4 Distribution of ideas' popularity when the social network is a circle

Consider a network in which agents are placed around a circle and have only one friend on each side. Because ideas diffuse face to face, the number of social network (directed) pairs with $j \in N_i \cap N_{i'}$ is simply $2(k_j - 1)$. It is also easy to see that there are only two directed pairs such that $j \in N_{i'}\backslash N_i$. In total, there are $2n$ directed pairs. Thus $Pr(j \in N_{i'}\backslash N_i) = \frac{2}{2n}$, so that $|N_{i'}\backslash N_i| = \frac{w}{n}$. The attachment kernel is then $A_t(k_j) = \frac{1-b}{bt}$, and the master equation for the steady state becomes $p(k) = (1 - b)p(k - 1)$. The first term is found to be $p_1 = b$, hence iterating the master equation gives the geometric distribution

$$p(k) = b(1 - b)^{k-1}.$$

However, when an idea is known n times, it cannot diffuse more. There are no bias as long as $k \leq n - 1$, but for $k > n$ it must be that $p(k) = 0$. For $k = n$ the master equation becomes

$$p(n) = \frac{(1 - b)}{b}p(n - 1) - 0$$

$$p(n) = \frac{(1 - b)}{b}b(1 - b)^{n-2} = (1 - b)^{n-1}.$$

The key point in the derivation above is that $Pr(j \in N_{i'}\backslash N_i)$ is independent of k_j. As long as this is the case, the same distribution will be obtained, because of the normalization by the sum (the competition among ideas).

Chapter 5

Knowledge diffusion and the structure of citation networks

Abstract

The distribution of citations received by scientific publications can be approximated by a power law, a finding that has been explained by "cumulative advantage". This chapter argues that socially embedded learning is a plausible mechanism behind cumulative advantage. More generally, irrespective of the source of cumulative advantage, the citation distribution reflects the popularity distribution of ideas, that is, "who knows what" determines "who cites what". Hence the shape of the citation distribution depends on the relative amounts of innovation and diffusion, with more diffusion (less innovation) implying flatter power laws. The structure of the social network on which diffusion takes place also influence the distribution of papers' popularity and, hence, citations received.

5.1 Introduction

In a paradigmatic model of the evolution of science, de Solla Price (1976) assumed that new papers tend to cite preferentially the papers already well cited. But do scientists cite the most cited papers or do they read the most well known ones? This chapter argues that the former is an artefact of the latter. In other words, the so-called preferential attachment of new papers to well-cited existing papers comes from a *preferential learning* mechanism through which scientists learn preferentially the popular papers and cite uniformly at random among the papers they know. While generally the literature claims that preferential citing occurs because scientists cite papers found in the bibliographies of other papers ("copying" models), this paper suggests that the main mechanism behind preferential learning resides

in the fact that scientists learn papers known by their peers. It is found that these assumptions predict a distribution of citations received similar to models of network growth by copying, and, to a good extent, compatible with recent empirical studies. In copying models, the distribution is determined by the relative amounts of papers found at random or through search. In the model introduced here, the citation distribution is determined by the way in which diffusion takes place, and by the way in which attention is allocated between knowledge consumption and knowledge production. Roughly speaking, preferential learning (learning popular ideas) is responsible for the power law distribution of citations, and the slope of this power law is determined by the relative amounts of time devoted to innovation (writing papers) or to diffusion (reading papers).

Since the prominent study of de Solla Price (1965), a large literature has repeatedly found that the number of citations received by scientific papers is approximately distributed according to a power law,[1] that is, the share of scientific papers having received k citations is $p(k) = Ck^{-\gamma}$, where C is a normalizing constant and with the power law exponent γ generally lying between 2 and 5. This finding has been explained by a mechanism that was originally pointed out to explain another well known power law in scientometrics called Lotka's law of scientific productivity, which states that the number of authors having published k papers follows again a power law. Simon (1955) proposed a model to derive Lotka's law as follows. He assumed that at each period one new paper appears, and it has a new author with probability v, otherwise this new paper is authored by an author who is chosen from the pool of existing authors with probability (called attachment kernel) proportional to his publication record, thus creating a "rich get richer" effect. These two principles (growth and positive feedback) are enough to generate Lotka's law. de Solla Price (1976), however, was interested in the *citation* network. Arguing for the existence of what he dubbed "cumulative advantage", de Solla Price derived a power law distribution of citations received from the following assumption: a paper receives citations at a rate proportional to the number of citations that it already has, that is, there is a process of preferential citing by which papers cite preferentially the papers that have been already well cited. The purpose of this chapter is to show that a power law-like distribution can emerge if we assume that scientists *cite uniformly at random*, but *learn preferentially* the papers which are already well known. A formal model is built using a growing two-mode (author-paper) network in which there is a link between an author and a paper if the author has produced or has learned the paper. Thus, the focus shifts from the dynamics of the epistemic

---

[1]Or a relatively similar law (to be discussed below). In all cases, the distribution is very (right) skewed, indicating a high level of inequality, where very few papers receive a very high number of citations, while the large majority of them receive none or a few.

network (the citation network alone) towards the co-evolution of the socio-epistemic network (the social network of authors, the citation network, and the author-paper network). In the model, the number of citations received by a paper is a direct function of its popularity, defined as the number of people knowing the paper. In fact, it seems reasonable to say that authors cite what they know; but it implies that the one-mode network dynamics (citations) depends on the structure of the two-mode network (popularity).

One mechanism that can create cumulative advantage in popularity is socially embedded learning. To see this, observe that if an agent chooses uniformly at random a paper known by another agent, then papers are chosen with probability proportional to their popularity. By reference to the landmark model of scale-free networks (Barabási & Albert 1999), in which newborn agents choose preferentially highly connected friends ("preferential attachment"), I call "preferential learning" the preferential attachment of people to papers that occurs when agents are choosing random ideas of random friends. In this case, the mechanism generating preferential learning is a two-mode version of what has been proposed for one-mode networks, a form of copying or referral (Kleinberg et al. 1999, Vázquez 2003, Jackson & Rogers 2007).

The model assumes that a population of agents read papers of others and cite part of what they know when writing their own papers . Under certain conditions (infinite population and complete social network), socially embedded learning can generate cumulative advantage in diffusion. Under additional restrictive assumptions about the productivity of the time spent writing/reading papers (only one paper can be written or learned per period), it is possible to estimate empirically that the share of time spent writing papers is about a third, with two thirds of the time devoted to reading/learning activities.

However, the structure of the social network can affect the distribution of popularity (chapter 4 section 4.5.1). Furthermore, there may be other factors influencing the diffusion process. To show that the creation/diffusion trade-off determines the shape of citation distribution, I propose a model in which agents can learn and create several ideas per time period, and ideas diffuse with probability proportional to popularity plus a constant a, intended to capture all the factors preventing pure preferential learning, including clustered (with many short cycles) social networks, preference for original or novel ideas, media (recommender systems and search engine) bias toward non popular ideas, etc.

The chapter is structured as follows. Section 5.2 reviews the literature. Section 5.3 presents the model. Section 5.4 gives empirical observations. The last section concludes.

5.2   Related literature

The model is related to a long tradition of modeling science, and in particular citation networks (section 5.2.1). The availability of very large datasets has revived a debate regarding the best functional form for empirical citation distributions (section 5.2.2). Meanwhile, many have argued that two-mode networks are an insightful way of representing science (section 5.2.3). This modelling standpoint is easily backed up by social theories of knowledge such as actor-network theory (section 5.2.4).

5.2.1   Models of citation networks

The literature on citation networks has become prominent following the seminal contribution of de Solla Price (1965) who explicitly conceived scientific publications and their bibliographies as a citation network, and explored the power law nature of the citation distribution. Power laws are ubiquitous in natural and artificial systems, and in the 1950's they had already been found for city sizes by Zipf (1949) and for biological genera sizes by Yule (1925), which motivated the paper by Simon (1955). Simon's model is in fact a growing two-mode network, but de Solla Price (1976) showed that it could be used to analyse a growing citation network. He assumed that new papers constantly arrive and cite previous papers with probability proportional to their already accumulated stock of citations.

One problem with this heuristic is that since papers start their life with zero citations, they would have no chance to receive their first citation. In his original article, Price simply counts publication as the first citation, such that everything happens as if papers were arriving with an in-degree equal to one. Dorogovtsev et al. (2000) have generalized de Solla Price's (1976) model[2] to allow for an initial attractiveness (call it a), and showed that the exponent of the (shifted) power law is given by $2 + a/h$. Hence, in this theory, the slope of the shifted power law depends on two parameters: the bibliography size (h) and the initial attractiveness (a).

The scale-free network model (Barabási & Albert 1999) has triggered an important amount of research on growing networks, providing new inputs for the understanding of citation networks. For instance, superlinear preferential attachment, the assumption that existing nodes receive new links at a rate more than proportional to their degree (Krapivsky et al. 2000), and initial attractiveness (Dorogovtsev et al.

---

[2]Newman (2009) showed that the in-degree distribution of Dorogovtsev & Mendes's (2000) model can be simplified as a ratio of beta functions, and Golosovsky & Solomon (2012) found its name to be the Waring distribution (this can be checked from its generating function defined in Irwin (1963), appendix II p.29). Throughout the paper I will refer to "Price's model", but credit is due to all these authors for extending and clarifying Price's original contribution.

2000) have been used to explain the in-degree distribution of the patent citation network by Sanditov (2005). Superlinear preferential attachment and aging (Dorogovtsev & Mendes 2000) have been used in the same context by Csárdi et al. (2007) and Valverde et al. (2007).

Another strand of literature has focused on the idea that authors typically cite recent literature, and then cite papers cited by this first article. An ingenious empirical study is that of Simkin & Roychowdhury (2012), who have shown that misprints in citations tend to repeat (in fact, the number of misprint repetitions follows a power law). They propose a model which allows them to estimate that 70 to 90% percent of the citations are copy/pasted from other papers.[3] Another recent contribution, by Peterson et al. (2010), explicitly distinguishes between two mechanisms underlying citations. In the *direct mechanism*, which occurs with probability c, a paper simply gets to be known and cited (the new paper cites every old paper with equal probability). In the *indirect mechanism*, which occurs with complementary probability $(1-c)$, the paper cited is taken from the bibliography of the known paper (thus the new paper cites every old paper with probability proportional to the in-degree of the old paper). These assumptions generate a rule for citation similar to preferential attachment with initial attractiveness. The latter comes from the random choice of papers, and preferential attachment comes from the search mechanism. They find a power law exponent equal to $1 + \frac{1}{c}$.

The difference between the two mechanisms described above is related to Jackson & Rogers's (2007) model for social networks, where newborn agents choose to link to random existing agents ("random meetings"), and to random neighbors of these first chosen agents ("search"). Atalay (2013) argued that these simple "mixed-attachment" (random and preferential) models cannot account for differences in citations received by papers of similar age. He modified the model by introducing a fitness function, as in Bianconi & Barabási (2001), to account for the differentiated quality levels of the papers.

In the context of patents, Ghiglino & Kuschy (2011) have proposed that the origin of the in-degree distribution is to be found in the heterogeneous nature of patent's applicability. They assume that there is a fixed number of patent classes, each of which sees new patents arriving following a Poisson process. Each patent i has a type μ_i and a broadness a_i which determines its support $F_i = [\mu_i \pm a_i]$.

---

[3]According to them, copying a misprinted citation constitutes evidence that the faulty author has not read the paper. One may dispute this interpretation, but their data at least suggests that sometimes authors know, among their references, which one cites another, and use this information to save time in preparing their reference list. From this, it is tempting to infer that an author became aware of a certain paper by seeing it referenced in another paper. Nevertheless, one can think of a situation in which an author knows a paper, see it subsequently cited by another paper, and copy the citation from this last paper.

When a new patent arrives, it cites the youngest patents which are in the support of technologically related but not-already-existing patents. As a result, the probability of patent i being cited is proportional to its broadness, a_i, so the distribution of citations received also depends on the distribution of the a_i's.

Bramoullé et al. (2012) also propose a model to explain the pattern of citations, though their main focus is on "long-run integration". Agents are of different types, and long-run integration occurs when the distribution of types of a node's friends becomes the same as in the whole population. Again based on Jackson & Rogers's (2007) idea of an initial random choice followed by neighbor-picking, Bramoullé et al. (2012) also add paper's type, and a homophilous process: the initial random choice of papers is biased towards same type (same subfield) papers. Their key insight is that, as time passes, the search process becomes more prevalent than the random meetings. Since in-degree is correlated with time, if the search process is unbiased we should observe a form of integration, which they confirm empirically on the American Institute of Physics dataset (1985-2003, 207,912 papers): as in-degree increases, the share of citations from same-type papers decreases.

This literature recognizes that authors must know the existence of a paper before they cite it, but does not explicitly keep track of who knows what. An exception is the TARL (Topic, Aging and Recursive Linking) model of Börner et al. (2004). They proposed that authors and papers co-evolve, but additional features of their model (papers' topics and aging) prevent them from finding closed-form expressions for the degree distributions. Notwithstanding details, the main difference with the model presented here is that they assume that authors, at each period, choose a number of items to read among all existing papers, and then scan the bibliographies of these papers to find additional literature. By contrast, in the learning-based model below, authors do not find new references by looking in the reference list of existing papers; they discover new papers to learn by choosing at random a paper known by a friend. Another important difference is that the TARL formulation does not explicitly model the trade-off between learning existing papers and producing new ones. This distinction is central here, where it is shown that the allocation of attention between knowledge consumption and production determines the shape of the citations distribution.

5.2.2   Degree distribution of citation networks

The statistical analysis of power laws is a rather technical and controversial topic (Perline 2005, Stumpf & Porter 2012) summarized in Clauset et al. (2009) and Gabaix (2009). Regarding citations data, the early contribution of de Solla Price (1965) and Seglen (1992) have been followed since the turn of the century by a

number of studies on large datasets, following the influential note of Redner (1998), who estimated that the tail of the distribution for the two datasets he studied could roughly be fitted by a power law with an exponent around 3. Arguing that a stretched exponential function (Laherrere & Sornette 1998) is helpful in fitting the left side (papers with few citations) but fails in the tail, Tsallis & De Albuquerque (2000) introduced a form which can fit the whole distribution, namely $p(k) \propto [1 + (g-1)\lambda k]^{\frac{-g}{g-1}}$, where g and λ are parameters, and can be rewritten $p(k) \propto (\frac{1}{(g-1)\lambda} + k)^{\frac{-g}{g-1}}$ to show that it is a particular case of the shifted power law $p(k) \propto (r+k)^{-\gamma}$, with a certain correlation of its parameters.

Several recent papers have used massive datasets to discriminate between the three major candidates: power law, shifted power law, and log-normal.[4] Stringer et al. (2010), looking at the ultimate number of citations received (i.e considering only papers which are not cited anymore), found that the distribution is a discrete version of the log-normal. Albarrán & Ruiz-Castillo (2011), with a huge dataset and considering citations received after 5 years, found that in most fields it is not possible to reject the hypothesis of a power law for the tails. Eom & Fortunato (2011), using a citation network constructed from papers published in American Physical Society journals, compared the fits given by the power law, the log-normal, and the shifted power law, finding that the last one is the best.

5.2.3 Two-mode networks

Sociologists have long been using two-mode networks to analyze social phenomena (Freeman 2003). For instance, the classical paper of Breiger (1974) studies the dual relationship between individual people and groups using a two-mode network where one mode is the actors, and the other is the groups to which they belong or the events that they attend. The intuition here is that actors are linked through their common affiliations or participation in certain activities. Indeed, although many classical concepts for one-mode networks have been extended to two-mode networks (Faust 1997, Latapy et al. 2008), two-mode networks are often projected into one-mode networks, in which two actors are linked if they have a common affiliation. However, many have argued that in most applications it is better to stick with two-mode networks, which allow, *inter alia*, to study the co-evolution of the two sets of nodes. For instance, Roth & Cointet (2010) studied empirically a scientific collaboration network and the concepts used in the associated articles, effectively characterizing socio-semantic co-evolutionary dynamics.

The model below is close to this literature as it adapts de Solla Price's (1976) and

---

[4]Or other more complicated distributions such as the modified bessel function, the stretched exponential, etc. See the literature review in Eom & Fortunato (2011).

Barabási & Albert's (1999) models to a two-mode network framework, following the work of Ramasco et al. (2004) which focused on co-authorship networks. Ramasco et al. (2004) considered both the set of actors and the set of papers, but only the *production* of papers. They proposed that at each time step, a new paper arrives, this paper having n_a authors. The authors are partly new, and partly taken from the existing set of authors proportionately to their existing number of authored papers. Thus, a difference with the model proposed below is that actors create new papers but they never learn existing ones. Another difference is that Ramasco et al. (2004) are interested in the distribution of the number of papers per author (whereas the interest here is in the number of agents knowing a given paper, which is the other side of the two-mode network), and on the projection of the two-mode network into its "co-authorship" one-mode projection (whereas projections are not considered here).

In summary, Simon showed that growth of the number of papers and cumulative advantage of authors regarding new authorship generates Lotka's law. de Solla Price (1976) showed that growth of the number of papers and cumulative advantage of papers regarding new citations generates scale-free citations networks.[5] Ramasco et al. (2004) showed that growth of the number of authors and papers and the cumulative advantage of authors regarding authorship generates Lotka's law and a scale-free co-authorship network. Chapter 4 shows that growth of the number of papers and cumulative advantage of papers regarding new diffusion generate a scale-free distribution of papers' popularity. This chapter will use this last result to study cumulative advantage in receiving citations.

5.2.4   Science in the making: some arguments for actor-paper systems with learning

Perhaps one of the most general, least controversial statement about scientists is that they are learning. But what do they learn? Once we assume that they learn knowledge items nicely encoded into "papers", the question becomes: what is the structure of this actor-paper network? There has been a number of models of knowledge production and diffusion (Cowan 2005), showing that network structure influences the distribution of knowledge. The point made here is that general information about knowledge diffusion can be inferred from citation networks. It is proposed to model the self-organization of science as a morphogenesis of the actor-paper system driven by heuristics: agents learn ideas, for instance from their friends, and, when they write a paper, they cite what they have learned before.

---

[5] "Scale-free" is used loosely here and in the literature in general. In fact Price's model features a shifted power law degree distribution, which is not scale-free except for (very) large degrees.

This approach, which significantly complicates the one-mode approach to citation networks, is justified by a large literature in science studies which has highlighted that many "social" processes affect the self-organization of science (Merton 1968, Latour & Woolgar 1979). Furthermore, not only are there arguments to introduce humans in citation network models, but there are also arguments to introduce non-humans in social network models of science. Proponents of actor-network theory (Star & Griesemer 1989, Latour 2005) contend that artefacts – material or not, but "non-humans" – indeed populate the net in which human agency takes place. One may argue that the most important such objects in science are the publications. Hence, one can choose as a fundamental unit for the description of science the relationship between an actor and a piece of knowledge. The dynamics of the system follows from the constitution of this relation (the learning event), because it is not independent of the rest of the network: it is this feedback loop which leads to self-organization. A proposal of formal modelling of these ideas appears in the next section.

5.3   The model

The objective of the model is to understand how diffusion impacts the structure of citation networks. The idea is to model the evolution of papers' popularity, and use this to derive the distribution of citations. I derive approximate results using the mean-field continuous framework[6] with additional assumptions about time and population size.

Denote by $k_j(t)$ the number of agents who know paper j (its popularity or diffusion). Consider a very large population of agents in a finite time horizon.[7] At each period, a randomly chosen (r.c.) agent takes a decision. With fixed probability b, she invests her time in writing λ_P papers. Otherwise, she learns λ_L papers that she did not previously know. I assume that a paper is chosen with probability proportional to $k_j + a$. In chapter 4, it is shown that, for large population and finite time, if agents are choosing a random idea of a random friend in a full social network, then diffusion is proportional to k_j. However, if the social network is a lattice with agents organized in a circle and with few links on each side, then the dependence on k_j disappears. Hence we can think of a as reflecting a departure from preferential learning, possibly due to the structure of the social network (see section 4.5.1).

---

[6]For the distribution of popularity, it would be possible to use more precise (chapter 4) or even exact methods (Bollobás et al. 2001).

[7]This is to ensure that finite-size effects are negligible.

Given these assumptions, the expected evolution of k_j is

$$\frac{dk_j(t)}{dt} = (1-b)\lambda_L \frac{k_j(t)+a}{\sum_{j=1}^{w_t}[k_j(t)+a]}. \tag{5.1}$$

The total popularity is equal to $t(b\lambda_P + (1-b)\lambda_L)$ so the sum in the denominator is equal to $t[(1+a)b\lambda_P + (1-b)\lambda_L]$, using $w_t = b\lambda_P t$. Defining

$$\xi \equiv \frac{b\lambda_P}{(1-b)\lambda_L}, \tag{5.2}$$

equation 5.1 becomes

$$\frac{dk_j(t)}{dt} = \frac{k_j(t)+a}{t(1+(1+a)\xi)}. \tag{5.3}$$

Using the initial condition $k_j(t_j) = 1$ (j is invented by one agent, at some time t_j), this differential equation has solution

$$k_j(t) = -a + (1+a)\left(\frac{t}{t_j}\right)^{\frac{1}{1+(1+a)\xi}}. \tag{5.4}$$

This leads to the probability distribution of ideas' popularity,

$$p(k) \propto (k+a)^{-2-(1+a)\xi}. \tag{5.5}$$

This distribution is a shifted power law. The constant creates a "shoulder" on the left of the distribution, but does not change the tails much since it converges to the classical power law for large k (i.e. for $k \gg a$). Equation 5.5 shows how the innovation/diffusion trade-off determines the shape of the popularity distribution: the more innovation (higher ξ), the more papers known only a few times. The more diffusion (lower ξ), the more papers known many times.

By knowing how diffused papers are, it is possible to determine how much they are cited. By assumption, when writing a paper, the author will cite some papers chosen at random among what she knows. Note that since an agent is chosen with probability $1/n$, and get to know λ_L papers with probability b and λ_P papers with probability $1-b$, on average an agent knows $t(b\lambda_P + (1-b)\lambda_L)/n$ papers. Moreover, the probability that paper j is known by a r.c. agent i is k_j/n. Hence,[8] the total

---

[8]As is usual in the literature, I do not explicitly take into account that a paper is never cited more than once by a newly arriving paper (a draw without replacement is approximated by one with replacement). This is reasonable because the probability that a paper would be chosen twice in a row tends to zero as the number of papers in the network tends to infinity.

number of citations received by paper j up to time t, denoted by $k_j^C(t)$, evolves as

$$
\begin{aligned}
\frac{dk_j^C(t)}{dt} &= b\lambda_P h Pr(\text{j chosen}|\text{j known}) Pr(\text{j known}) \\
&= b\lambda_P h \frac{n}{t(b\lambda_P + (1-b)\lambda_L)} \frac{k_j}{n} = \frac{hk_j}{t(1+1/\xi)}
\end{aligned}
\tag{5.6}
$$

using equation 5.2. Since $k_j(t)$ is given by equation 4.17, it can be substituted into equation 5.6. However, to derive the citation distribution, $k_j(t)$ might be simplified by noting that for large t, the term $-a$ will be negligible so that

$$
k_j(t) \approx (1+a) \left(\frac{t}{t_j} \right)^{\frac{1}{1+(1+a)\xi}}.
\tag{5.7}
$$

Note that using (5.7) instead of (5.4) would give a *pure* (Pareto) instead of a *shifted* power law for the popularity distribution, but with same power law exponent. This is not surprising since power law behavior is observed for large k, which is the assumption made to simplify (5.4) to (5.7). Under this approximation, one can use (5.7) into (5.6) to get

$$
\frac{dk_j^C(t)}{dt} = \frac{h(1+a)}{t(1+1/\xi)} \left(\frac{t}{t_j} \right)^{\frac{1}{1+(1+a)\xi}},
\tag{5.8}
$$

which, with the initial condition $k_j^C(t) = 0$, has solution

$$
k_j^C(t) = (1+a)h \frac{\xi(1+(1+a)\xi)}{1+\xi} \left(\left(\frac{t}{t_j} \right)^{\frac{1}{1+(1+a)\xi}} - 1 \right).
\tag{5.9}
$$

Using the cumulative distribution function and assuming that the t_j's are uniformly distributed (see section 3.1), equation 5.9 leads to the citation distribution

$$
p^C(k^C) \propto \left(k^C + (1+a)h \frac{\xi(1+(1+a)\xi)}{1+\xi} \right)^{-2-(1+a)\xi}.
\tag{5.10}
$$

This is again a shifted power law. Equation 5.10 shows the effect of the innovation/diffusion trade-off: more innovation (higher ξ) makes the power law steeper, and more diffusion (lower ξ) makes it flatter. The parameter a, which captures the details of the diffusion process, has an equivalent effect. Hence, more uniformly random learning (that is, less preferential or higher a) has an effect similar to more innovation (less diffusion). It should also be noted that the exponent is the same as that of the distribution of papers' popularity (equation 5.5), and is independent of the bibliography size h. The key point is that the citation distribution reflects the

popularity distribution, and the popularity distribution is determined by the learning process and the innovation-imitation trade-off. Hence, in contrast to existing models where the power law exponent of a citation distribution is explained by the prevalence of random and preferential citing, this model explains it by the prevalence of random and preferential *learning* (not citing), *and* the innovation-imitation trade-off (ξ).

5.4   Empirical observations

The model predicts[9] (equation 5.10) that the citation distribution is a shifted power law

$$p^C(k^C) \propto (k^C + r)^{-\gamma}, \tag{5.11}$$

with

$$\gamma = 2 + (1 + a)\xi \tag{5.12}$$

and

$$r = (1 + a)h\frac{\xi(1 + (1 + a)\xi)}{1 + \xi}. \tag{5.13}$$

There are too many parameters in the model as compared to the number of parameters that can be estimated from data on the citation distribution (γ and r). In other words, the model is under-identified. The parameters of the model are the probability of innovation, b, the productivity of learning (λ_L) and innovating (λ_P), the degree of non-preferential learning, a, and the average bibliography size, h, which also gives the mean of the distribution of citations received. The symbol ξ, which is not a primitive parameter, summarizes the innovation-imitation trade-off in terms of b, λ_L and λ_P (equation 5.2). To obtain identification, let me assume $\lambda_P = \lambda_L = 1$[10]. If one further assumes $a = 0$, then

$$\gamma = 2 + \frac{b}{1 - b} \tag{5.14}$$

and

$$r = \frac{b}{1 - b}h. \tag{5.15}$$

Equations 5.14 and 5.15 form a system of two equations with two unknowns, b

---

[9]It should be emphasized that the estimation of the parameters relies on the approximations made to obtain analytical results. Ideally, one would fit a full-fledged agent-based model, with specific network structure, number of agents, etc.

[10]It is possible to estimate the more general ξ instead of b. Here I choose to present the estimates of b because it has a direct interpretation as a probability, which is more intuitive. However, it should be remembered that what we are really measuring is the innovation-diffusion trade-off in a general sense. From the estimates of b presented in table 5.1, ξ can be calculated using $\xi = b/(1 - b)$.

and h. Since h can be observed independently (as the mean in or out-degree of the network), it will provide a useful consistency check.

To estimate the parameters, I propose to use a Waring distribution (Irwin 1963)

$$p(k^C) = \frac{B(k^C + r, \gamma)}{B(r, \gamma - 1)} \tag{5.16}$$

instead of the shifted power law, mostly because the Waring is a discrete distribution, as are the data. This equation can be seen as a discrete form of the continuous function (5.11). In fact, many related (i.e. linear) models of network growth (de Solla Price 1976, Dorogovtsev et al. 2000, Jackson & Rogers 2007, Peterson et al. 2010) have an exact solution of the form of (5.16), and an approximate solution of the form of (5.11). Therefore, it seems more legitimate to use (5.16) than a discretized shifted power law.

By assuming that the data are distributed according to a Waring, it is possible to obtain the empirical values of the parameters, r and γ, by maximum likelihood. Here I study the time evolution of these parameters in two citation networks of high energy physics papers. These two publicly available datasets[11] are citation networks of a subset of *arXiv* publications: High Energy Physics Theory (HepTh) and High Energy Physics Phenomenology (HepPh).[12] They cover almost completely the history of these subsets of *arXiv*, until 2003. In both cases, a citation network is constructed for each year, where year T network includes all the papers published on *arXiv* before December 31st of year T.[13] Descriptive statistics can be found in the leading columns of table 5.1. In practice, a Nelder-Mead algorithm was used to maximize the logarithm of the likelihood function associated to equation 5.16. For comparison purposes, the fit of a power law is also added.[14] The sixth and seventh columns of table 5.1 show the estimated parameters assuming a power law (for the tail), and the eighth and ninth columns give the estimated parameters assuming a shifted power law (for the whole distribution). Early networks (1992-93) are not reported. The resulting look of the predicted laws can be seen in figure 5.1. The fits are reasonably good, although a deviation can be observed when the complementary cumulative distribution is plotted, especially for the HepPh network (subpanel c of the right panel of figure 5.1). Second, one sees that the predicted mean in-degree is

---

[11]Downloaded from http://snap.stanford.edu/data/index.html#citnets. Credit is due to Gehrke et al. (2003) and Leskovec et al. (2005).

[12]There is an overlap between the two. *arXiv* is a server where scientists, mainly physicists, can freely post pre- and post-prints of their work. See *arXiv.org*.

[13]The model presented here, as Price's model, predict the distribution of all papers and all citations to those papers. An ideal test of these models would require data on science as a whole.

[14]The power law fit was done using the procedure and code of Clauset et al. (2009) (translated to R by Laurent Dubroca)
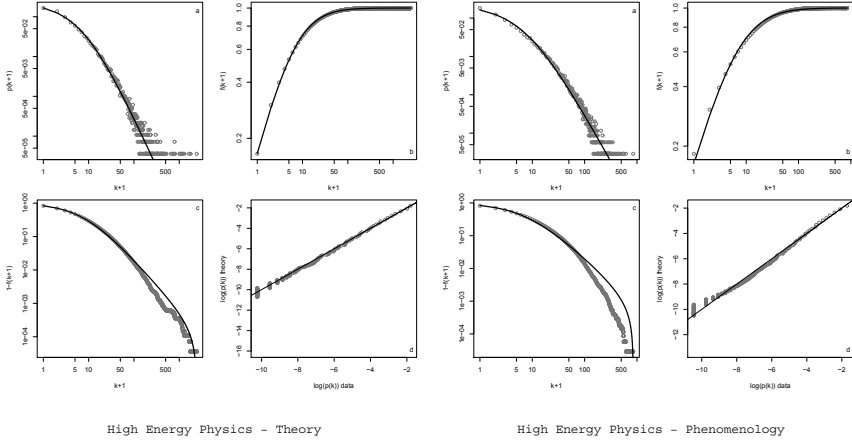
Figure 5.1: Maximum likelihood fit for the citation distributions of the Hep-Th (left) and Hep-Ph (right) networks at the end of 2003. In panels a, b and c, the grey points are empirical values, and the black lines are the theoretical results constructed by using equation 5.16 with values of the fitted parameters (see tables 5.1) from $k = 0$ to the maximum observed degree, and renormalizing such that it sums up to one. All subpanels of a given panel display the same raw data. a) Probability distribution. b) Cumulative distribution. c) Complementary cumulative distribution. Panel d) is a Quantile-Quantile plot, where the black line materializes perfect fit.

close to the observed one (columns 4 and 5 of table 5.1). Finally, the last column gives the corresponding value of b, computed using the estimated shifted power law exponents and equation 5.14. The shifted power law exponent is strikingly stable over time and across the two networks, and imply $b \approx 1/3$. If one takes this exercise seriously, it says that high energy physicists spend a third of their time working on their own papers and two thirds learning the work of others.[15]

---

[15]A lower estimate results if we estimate h directly as the mean degree. In this case. one can solve the system 5.12-5.13 for a and b in terms of h, r and γ. The estimated values of a are generally small, but this new estimation can slightly affect the values of b. For the last year of each network, I find $(a = 0.78, \xi = 0.29)$ for the Hep-Th, and $(a = 2.10, \xi = 0.16)$ for Hep-Ph. Assuming $\lambda_P = \lambda_L = 1$, we have $b = \xi/(\xi + 1)$ equals to 0.22 in the Hep-Th network and 0.14 in the Hep-Ph network.

1	2	3	4	5	6	7	8	9	10
Year	Number of edges	Number of nodes	Empirical in/out degree	Predicted in/out degree	Minimum degree for PL fit	PL exponent	"Shoulder" param. for shifted PL	Shifted PL exponent	Estimated innovation probability
T	w_t		$\langle k \rangle$	h	k_{min}	γ_{PL}	r	γ_{SPL}	b
High Energy Physics - Theory									
1994	12,879	4,924	2.62	2.56	6	2.53	2.79	2.99	0.50
1995	28,131	7,078	3.97	4.01	24	3.11	3.65	2.85	0.46
1996	53,091	9,541	5.56	5.63	25	2.79	4.37	2.72	0.42
1997	87,713	12,124	7.23	7.33	28	2.63	4.80	2.59	0.37
1998	125,163	14,766	8.48	8.65	29	2.57	5.11	2.52	0.34
1999	167,110	17,485	9.56	9.79	36	2.61	5.66	2.52	0.34
2000	217,790	20,523	10.61	11.09	58	2.77	6.27	2.51	0.34
2001	271,863	23,587	11.53	12.15	100	2.98	6.90	2.52	0.34
2002	333,973	26,792	12.47	13.28	62	2.75	7.53	2.52	0.34
2003	352,807	27,770	12.70	13.60	55	2.70	7.60	2.51	0.34
High Energy Physics - Phenomenology									
1994	11,387	4,423	2.57	2.53	22	3.56	1.68	2.51	0.34
1995	29,802	7,276	4.10	4.10	20	3.04	3.05	2.62	0.38
1996	58,904	10,478	5.62	5.65	16	2.74	4.07	2.59	0.37
1997	98,307	14,013	7.02	7.08	25	2.94	5.19	2.61	0.38
1998	142,934	17,736	8.06	8.23	76	3.74	5.82	2.59	0.37
1999	201,300	21,739	9.26	9.51	56	3.37	6.34	2.55	0.35
2000	263,313	25,764	10.22	10.57	64	3.38	6.75	2.52	0.34
2001	333,942	29,877	11.18	11.86	72	3.40	7.21	2.51	0.34
2002	411,527	34,014	12.10	13.07	94	3.49	7.66	2.50	0.33
2003	421,578	34,546	12.20	13.22	94	3.47	7.69	2.49	0.33

Table 5.1: High Energy Physics Theory and Phenomenology citation network statistics. See the main text for details.

Study	Dataset	γ	r	h	b
Tsallis & De Albuquerque	ISI cohort 1981	2.89	17.2	19.3	0.47
Tsallis & De Albuquerque	Phys. Rev. D	2.56	12.0	21.5	0.36
Jackson & Rogers	Small world	2.63	3.1	5.0	0.39
Newman	Net. science	2.28	6.4	22.8	0.22
Peterson et al.	ISI cohort 1981	3.20	25.0	20.8	0.55
Peterson et al.	Phys. Rev. D	3.10	31.0	28.2	0.52
Peterson et al.	Chemists	2.94	37.0	39.6	0.48
Golosovsky & Solomon	Physics Journals	3.15	27.5	23.9	0.53
Golosovsky & Solomon	Physical Review	3.15	10.2	8.87	0.53
Eom & Fortunato	Am. Phys. Soc. 1950	5.60	7.92	2.2	0.78
Eom & Fortunato	Am. Phys. Soc. 2008	3.10	9.90	9.0	0.52

Table 5.2: Estimates from the literature. These authors estimated slightly different distributions (shifted power law, truncated shifted power law, Waring distribution, Tsallis distribution), and using different methods, so the results are not perfectly comparable. The values of h and b are deduced from the values of γ and r (using $E[p(k)] = h = r/(\gamma - 2)$ and (5.14)). In some cases the value of r is deduced from the reported h.

Furthermore, I collected estimates of γ and r in the literature (table 5.2). Different authors have used different datasets and estimation techniques, so that the results can vary significantly among studies ($b \in [0.2, 0.8]$). Since most studies in table 5.2 deal with physics papers, there are no reasons to expect such a diversity. Besides data and estimation issues, the problem can be that the values of the parameters are changing as the network evolves, making networks of different size hard to compare.

In fact, the results in table 5.1 suggest that the networks are accelerating. Acceleration (Dorogovtsev & Mendes 2001, Cooper & Prałat 2010) or densification (Leskovec et al. 2005) means that the average degree is increasing, which is the case here (column 4). In a citation network, this can happen only if new papers have, on average, a larger bibliography than old papers. It is difficult here to say if the network is truly accelerating or if we observe simply a biased data truncation (the oldest papers make more citations to outside-sample papers). However, at a global level, there is evidence of an increase of bibliography size (Biglu 2008). This acceleration shows that knowledge systems are in disequilibrium, leading one to wonder what are the sources and consequences of this acceleration. If scientists have to cite more papers over time, how many will they have to cite in the future? If it increases unboundedly, it could be the case that writing one paper will require reading more and more papers. In fact, it seems reasonable to think that the more knowledge is already available, the more should be learned to push the frontier, a point emphasized by Jones (2009).

Finally, note that the model proposed here does not take into account several

features of real-world citation networks which have been shown to be relevant, such as aging (Dorogovtsev & Mendes 2000), burst dynamics (Ratkiewicz et al. 2010) and paper intrinsic value (Adamic & Huberman 2000, Bianconi & Barabási 2001, Atalay 2013).

5.5 Conclusion

The availability of large scale citation networks allows us to observe the structure of knowledge. The cumulative or sequential nature of knowledge reveals that it is created at least partly by recombining existing ideas. Therefore, innovation itself is conditioned on diffusion. Typically, authors cite papers that they have not written themselves. Since there must be both innovation and diffusion, the relative amount of each activity shapes the distribution of ideas' popularity, and therefore the distribution of citations received.

This chapter developed a quantitative model relating the innovation - diffusion trade-off and the type of learning dynamics (self-reinforced or not) to the parameters of the power law distribution of citation networks. Hence, it showed that the hypothesis of social diffusion (in an infinite population) explains the structure of citation networks equally well as the current dominant hypothesis of bibliographic (asocial) search. More generally, it is argued that learning preferentially well-known ideas ("preferential learning") naturally generates power law distribution of popularity, and hence of citations. Moreover, the shape of citation distributions (the value of the power law exponent) is determined by the innovation-imitation trade-off, which is absent from current models of citation networks. The more time is allocated to diffusion instead of innovation, and the more learning is preferential, the flatter (more unequal) we should expect citation distributions to be. As a result, it can also be concluded that the structure of the social network on which knowledge diffusion takes place influences the distribution of citations, with social networks conducive of preferential rather than random learning leading to more unequal citation distributions.

Chapter 6

The size of patent categories: USPTO 1976-2006

Abstract

Categorization is an important phenomenon in science and society, and classification systems reflect the mesoscale organization of knowledge. The Yule-Simon-Naranan model, which assumes exponential growth of the number of categories and exponential growth of individual categories predicts a power law (Pareto) size distribution, and a power law size-rank relation (Zipf's law). However, the size distribution of patent subclasses departs from a pure power law, and is shown to be closer to a shifted power law. At a higher aggregation level (patent classes), the rank-size relation deviates even more from a pure power law, and is shown to be closer to a generalized beta curve. These patterns can be explained by assuming a shifted exponential growth of individual categories to obtain a shifted power law size distribution (for subclasses), and by assuming an asymmetric logistic growth of the number of categories to obtain a generalized beta size-rank relationship (for classes). This may suggest a shift towards incremental more than radical innovation.

6.1 Introduction

Categorization is at the basis of reasoning. Theorizing about scientific and technological systems is no exception, and always relies on the grouping of several items into "categories". For instance, the concepts of paradigms, research fields, school of thought, epistemic communities, etc. are all based on the idea that an underlying grouping can be meaningfully established. Elements in these categories ("bio-technologies", "economics", "subclass N234", "keynesians") are then taken to behave in the same way. At the very least, the analyst can argue that elements

91

within a category have a degree of homogeneity which is much higher than elements taken from different categories. Since analyzing a number of categories is simpler than analyzing every single element, categorization reduces the dimension of the problem.

Categorization, therefore, is at the heart of thought processes. This implies that categories are not simply useful to describe reality, they are the main tool to *construct* it. Categories, when they are created as nouns, can have a predicate and become a subject. They enter discourses with their own identity, and shape our understanding of reality. Classification systems are essential tools in the creation of routinized habits of thoughts. Hence, when a classification is put to use, one may argue that it creates a feedback on the system it describes. Classification systems are institutions which often legitimate the items that they classify. This affects the future evolution of the items, and their relation (boundaries) with other items. Along this line of argument, the process of categorization is performative. The evolution of the technological classification system therefore provides data on how society understands its technological artefacts and legitimizes them through the process of categorization.

In this chapter, I propose an attempt at clarifying some of the key processes underlying the evolution of technological and scientific classification systems by studying in detail one of the most important, relatively well defined quantity: the size distribution of categories (or the size-rank relationship, which by construction is less noisy). I study the US patents granted by the USPTO between 1976 and 2006, partitioned at the level of more than 400 classes and 100,000 subclasses.

The size distribution of patent *sub*classes is well fitted by a shifted power law, in agreement with a slightly modified version of the Yule (1925)-Simon (1955)-Naranan (1970) models. However, at the level of *classes*, the size distribution is less skewed. The small sample (428 categories) suggests to study the size-rank relationship instead of the size distribution. The size-rank relationship, at the level of classes, is not a power law (Zipf's law). An exponential relation was recently proposed (Carnabuci 2013) for this data. Here I find that a generalized beta, suggested for size-rank relationships by Martínez-Mekler et al. (2009), fits the data better. I give a simple and original explanation for this fact, answering partially an open problem stated in Egghe (2012). The reason for the departure from Zipf's law is that the number of categories tends to grow faster in the beginning than in the end, as compared to the growth of individual categories. The model of Naranan (1970), which is a simplified version of Yule's (1925) and Simon's (1955) models, derives Zipf's law by assuming an exponential growth of both the number of categories and the number of items per category. To obtain a generalized beta (of

the first kind) for the size-rank relationship, I find that one should instead assume an asymmetric S-shaped curve for the number of categories.

The chapter is organized as follows. Section 6.2 gives a background discussion on category systems and technological change, and reviews existing literature. Section 6.3 describes the data and the methodology. Section 6.4 presents empirical results. Section 6.5 proposes theoretical models consistent with the observed empirical laws. Section 6.6 discusses the results. The last section concludes.

6.2 Literature review

The purpose of this section is to provide a general discussion of what categories mean, why it is important to study them, and how this general theoretical background applies to the case of technological categories. A review of the literature on the evolution of technological domains using patent categories follows.

6.2.1 Theoretical background

The philosophy of category systems[1] has traditionally distinguished between Aristotle realism (categories of things do exist) and Kant conceptualism (what really exists are the categories of understanding, based on experience). Husserl proposed an encompassing view where the two systems, categories of meaning and ontological categories, co-exist and are related. Foucault (1966) insisted on the idea that words, the categories making up discourses, are not descriptive tools but genuinely *construct* the world. Latour (2005) concluded that the social scientist should not overimpose her own categories over the actors she analyzes. Instead, the analyst should follow the actors, and see how they create categories themselves.

A more realist view is that of *natural kinds*. A natural kind is defined as one that "corresponds to a grouping or ordering that does not depend on humans". Chemistry is said to provide the least controversial example of natural kinds (the periodic elements table), whereas biological species classification are less easily taken for granted. From a metaphysical point of view, one is interested in the essence of natural kinds, that is, "the property or set of properties whose possession is a necessary and sufficient condition for a particular's being a member of a kind". The existence and relevance of these essences is the point of view of "essentialists". The view according to which there are "genuinely *natural* ways of classifying things" is called *naturalism.*

---

[1]see the Stanford Encyclopedia of Philosophy for *Categories* and *Natural kinds*, from where the quotes are taken (http://plato.stanford.edu, accessed 28/08/2014).

On the other hand, constructivists do not believe that classifications reflect the "real" world. In its weak version, constructivism does not deny the existence of natural kinds, but doubts that we can actually see them. Strong constructivism however rejects the mere existence of natural kinds. "Ontological relativism" can be defined as "the view that all entities, processes, relations and theoretical posits are relative to a certain conceptual scheme". More generally, constructivists argue that categories are created, constructed by the observer. In some cases, it is even argued that the objects – not only the categories as a concept, but the actual objects – are constructed.

Whatever the nature of categories – real or constructed – they are the key building blocks of discourses, including scientific discourses. More generally, a criterion for the existence of a category is that it is a level of aggregation at which a given law holds. According to the "cluster kind realists", "a natural kind is any (...) family of co-occuring properties that may be employed in inductive inference for the purpose of scientific explanation". Quine argued that "it is the similarity or sameness of kinds between instances that permits an induction". Putnam's (1975) theory of semantic externalism holds that the *meaning* of what people say is not in their head, but in the head of experts - the linguistic community- who collectively know what things are.

In recent years, there has been an expanding literature in sociology about classification, notably some historical studies of controversial classification systems, such as the classification of diseases (Bowker & Star 2000). Some analytical insights from this literature are worth mentioning (Shepherd 2010), and all have to do with the fact that classification exerts an effect on the users of the classification scheme, or even on the elements being classified. First, there is evidence that the degree of institutionalization of the classification system influences the perception of hybrid members (elements with multiple categories): once a classification system is well established, there is a penalty to hybridity. Second, classification essentially consists in drawing boundaries. Boundaries should be thought of as interfaces rather than borders (Bowker & Star 2000). How communication takes places within boundaries is different than how it takes place between; hence classification institutionalizes possible interactions. Third, the extent to which the users of classification systems can shape the classification changes the way in which users interpret and use the classification system. Finally, actors exercise power to obtain classification systems to their advantage.

6.2.2 The classification of technologies

Technologies at the mesoscale

For Arthur (2009), a technology is something that relies on the mastering of natural phenomena to produce a useful artefact. In his words, technology is "a collection of phenomena captured and put to use". Acknowledging that technology builds out of itself, we see that a technology is made of technologies, thus a technology is a "complex of interacting phenomena". Here is the central point of the discussion: since there are families of phenomena (chemical ones, electrical ones, quantum ones), there are families of technologies based on these phenomena. Arthur goes a step further by arguing that "each grouping forms a language within which particular technologies – particular devices and methods – are put together as expressions within that language". Hence, technologies form clusters, which he calls *domains*, because they are based on the same phenomena, or because of some other shared characteristics or purposes. Individual technologies and domains, though both hierarchical constructs (i.e. having their sub-technology or sub-domains), are distinct. To be sure,

> A technology (individual, that is) does a job; it achieves a purpose -
> – often a very particular purpose. A domain (technology-plural) does
> no job; it merely exists as a toolbox of useful components to be drawn
> from, a set of practices to be used. A technology defines a product, or a
> process. A domain defines no product; it forms a constellation of tech-
> nologies –a mutually supporting set – and when these are represented
> by the firms that produce them, it defines an industry. A technology is
> invented; it is put together by someone. A domain (...) is not invented;
> it emerges piece by piece from its individual parts. A technology – an
> individual computer, say – gives a certain potency to whoever possesses
> it. A domain –the digital technologies – gives potential to a whole econ-
> omy that can in time become transmuted into future wealth and political
> power.

However, it is not always obvious to attribute a unique category to a given innovation. From the inventor perspective, Arthur (2009) describes the process of choosing a category (a "palette of components") for a new device as *domaining*. Sometimes, this is automatic, sometimes more difficult. Often, if the technology is large enough, it will belong to several domains.

Rationalizing practice: classification as legitimation

If technologies can be meaningfully categorized, the information feedback provided by the category system will in turn influence further technological development. From a pragmatic stance, things are real if they are real in their consequences, so if firms use category systems to search for technologies and build their own, categories are ontological. But how exactly do these categories map with the human perceptions, or construction, of them? Nelson (2006) describes technological evolution as the co-evolution of a body of practice and a body of understanding. He describes the role of the body of understanding as one of "rationalizing" the practice.

> (...) what makes the evolution of human practice, and especially technology, different from the evolution of animal behavior as studied by ethologists is exactly that extant human practice is generally supported by a rather elaborate body of reasons, or rationalizations.(...) To the extent that technology is seen as not simply a body of practice, but also a body of understanding, the nature of the evaluation and selection processes becomes more complicated. While the criteria for selection on the former aspect may well 'fit' with user need, the criteria for the later may appear to be the 'ability to explain observed relevant facts and enable problems to be solved and progress made'. The selection processes and those who control them, as well as the criteria, may well be different. For practice, the process is ultimately under the control of users, or their agents; for understanding, the control rests with the community of technologists".

In this chapter, I consider categorization as a process of codification of an understanding concerning the technological system, and I argue that the dynamics of patent classes and subclasses constitute a window on the "community of technologists". Using mathematical and statistical modelling, it is possible to uncover the most fundamental principles at play in the growth of classification systems. Perhaps the most important of these principles is *creation*, i.e. the fact that *new* categories are created over time.

The emergence of new technology categories

Technological evolution is reflected in the evolution of the classification system and this suggests to study the dynamics of categories creation (Strumsky et al. 2012). Classification is a particularly important topic in innovation studies because by definition, innovation challenges existing schemas (Hicks 2011). Economists of

innovation have traditionally made a distinction between incremental and radical innovations. In the context of classification, the former can be interpreted as being a new technology which fits perfectly in the existing classification scheme. This sort of innovation falls into what Arthur (2009) calls "structural deepening", the addition of subsystems and subassemblies to an existing technology so as to improve basic performance, adapt to wider tasks, improve reliability and safety, or simply react to changed circumstances, for instance on the demand side. On the other hand, a radical innovation can be defined as a new technology which creates a new category (perhaps by creating a need to reframe the classification system, or simply by being *sui generis*, the first innovation of its own kind) – this would be called redomaining or revolution in Arthur's terms.

6.2.3 Growth and distribution of technological domains

The growth of technological domains has been deeply scrutinized in the economics of technical change and development (Schumpeter 1934, Dosi 1982, Pasinetti 1983, Pavitt 1984, Freeman & Soete 1997, Saviotti 1996, Malerba 2002). A recurring theme in this literature in the high heterogeneity among sectors. Besides heterogeneity due to the very nature of the underlying knowledge base, sectors are also at different stages of their life cycles (Vernon 1966, Klepper 1997).

Life-cycle theories of technological change suggest non linearities in the evolution of individual technological domains. Andersen (1999) fits logistic growth models, using USPTO data for the period 1890-1990, and aggregated at the level of 56 technological groups (collections of patent classes). Andersen (1999) does not fit logistic growth for the whole period (1890-1990) but instead for subperiods, arguing that a logistic growth episode constitutes one cycle. Hence, in this theory, an individual category follows repeated S-shaped patterns. In this paper, I do not look for detailed life-cycle patterns in the data. Instead, I seek for the most parsimonious model that gives reasonable aggregated results. Hence, the results are less precise, and less informative of individual sectors' histories. On the other hand, the models and results of this paper are more powerful in explaining why other, potentially very different datasets, exhibit similar regularities ("universal laws").

Similarly, Carnabuci (2013) detected a departure from pure multiplicative growth (Gibrat's model) for the size of individual categories, which I do not account for in the models of section 6.5. It turns out that at both the subclass and class levels, assuming exponential growth of individual categories gives reasonable frequency-size or rank–size distributions, if associated with an appropriate growth function for the number of categories.

Finally, using the almost complete record of US patents, Youn et al. (2014) found

that the number of subclasses follows the same trend as the number of patents until
about 1870, but grew less fast afterwards. Remarkably, however, the number of
unique *combinations* of subclasses used in a patent keeps the same trend as the
number of patents. The data used in the present study does not contain multiple
classifications, but only the main patent class/subclass, and covers only the period
1976-2006, as described in the next section.

6.3   Data and method

6.3.1   The United States Patent Classification System

Falasco (2002) describes three types of rationale behind the United States Patent
Classification System(USPCS). The classification by *industry* was the original method
when the first USPCS appeared in the beginning of the XIXth century. The classifi-
cation by *structure* ("arrangement of components") is sometimes useful, for instance
composite materials can be classified according to the arrangement of their parts.
The classification by *utility* (proximate function) was "adopted in the early XXth
century as the fundamental principle for classifying prior patent art in the USPCS".
By proximate function, it is meant the fundamental function of the invention, not
some example application in a particular device or industry. The classification by
utility includes also the classification by effect or product, which "provides collec-
tions based on *result*", for instance measuring, illuminating, etc.

The USPCS attributes to each patent at least one subject matter. A subject
matter includes a main class, delineating the main technology, and a subclass,
delineating processes, structural features and functional features. All classes and
most subclasses have a definition. All subclasses within a class are arranged in a
class schedule, "with the most complex and comprehensive subject matter generally
at the top of the schedule, and the least complex and comprehensive at the bottom"
(USPTO 2012). Subclasses have indentation levels. Primary subclasses are the
main type of subclass, and each patent must be assigned to at least one of them.
Some subclasses are called "Alpha subclasses". These were previously unofficial
subclasses, created by examiners to ease their work. They are identified by their
parent primary subclass, adding one or two letters. Finally, note that these are the
patent claims which are classified, and the patent inherits the classification of its
claims. The main classification is the one of its main claim.

Patent examiners follow a rather precise algorithm for finding the appropriate
subclass (USPTO 2012): starting at the top, scan downward looking only at main
line subclasses until one is found that provides for any portion of the claim. Within
this one, scan downward subclasses which are at one higher indentation level until

one is found that provides with at least a certain portion of the subject matter, and go ahead similarly to the next indentation level. When no subclasses can be found, stop the process and classify at the last subclass selected.[2]

Classification is used mainly for the search of prior art. It is also used to determine which department, that is, which examiners will evaluate the patent. It is also widely used by business analysts.

6.3.2   The NBER patent data (1976-2006)

I use the well-known NBER patent dataset (Hall et al. 2001), updated to 2006.[3] It contains information on 3,210,361 utility patents. 1335 patents have a "NA" class, corresponding either to a "withdrawn" subclass (1328) or to a "D" subclass (7 patents). I removed these 1335 patents, leaving a total of 3,209,056 uniquely allocated to 428 classes and 119833 subclasses ("primary" US class/subclass). Subclasses are detailed at the 6 digits level, and I consider alpha subclasses as distinct. Throughout, I use the current classification system, that is, the category of each patent is its category in 2008 ("CCL" variables in the NBER file, i.e. the variables "cclass" and "nclass").

6.4   Empirical observations

This section presents empirical results on the size distribution of subclasses, and the size-rank relation for classes. The size of a category is simply defined as the number of patents within it.

6.4.1   The size distribution of subclasses

The size-distribution of patent subclasses is a heavy tail distribution, and it is relatively well behaved (not too noisy). After trying a number of classical candidates (power law, stretched exponential, Weibull, etc...), two distributions, the Waring (a discrete version of the shifted power law) and the lognormal, were found to give a good fit. The parameters of these two distributions were estimated using maximum likelihood. A modified Waring distribution was used, defined so that it is properly

---

[2]It is interesting to note that this is a case of a formal, codified rule that should be followed to produce additional codified knowledge (a classification). It highlights that classification is done by following heuristics, as would be expected from bounded rationality and evolutionary economics theory (Simon 1947, Nelson & Winter 1982). Note that the knowledge of this procedure can in principle be exploited to understand the growth of subclasses as a function of their indentation level and position in the class schedule.

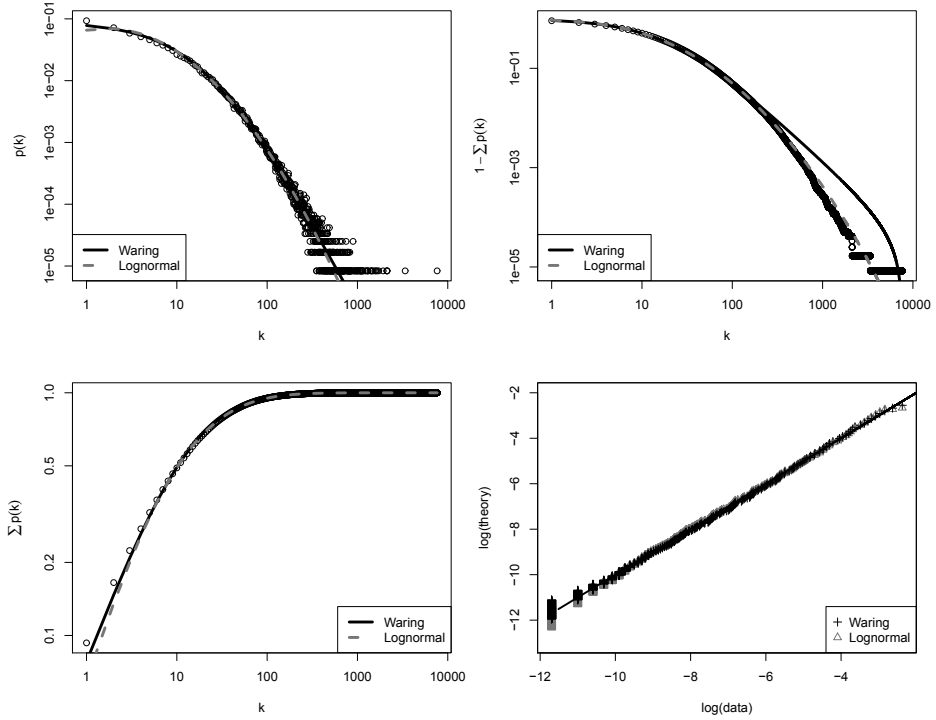[3]see https://sites.google.com/site/patentdataproject/Home. I used the file pat76_06_assg.dta.

Figure 6.1: Size distribution of patent subclasses. Top left: Probability mass function. Top right: Complementary cumulative distribution. Bottom right: Cumulative distribution. Bottom right: QQ-plot on double log axis. The fitted curves are a Waring and a log-normal distribution.

normalized over the range $k = 1 \ldots \infty$ (instead of $0 \ldots \infty$):[4]

$$p(k) = \frac{d + \gamma - 1}{d} \frac{\mathrm{B}(k + d, \gamma)}{\mathrm{B}(d, \gamma - 1)}. \tag{6.1}$$

The parameters and R^2 were computed for the cumulated system (all patents granted up to a given year), at 6 periods of 5 years interval (table 6.1). The visual fits for the system as of 2006 are provided in figure 6.1. Although this fit is visually good, the log-transformed data does not pass the Anderson-Darling test for normality. The values of the likelihood (not reported) were slightly but systematically higher for the Waring distribution than for the lognormal.

---

[4]The normalization constant $C = \frac{d + \gamma - 1}{dB(d, \gamma - 1)}$ is found as follow. First impose $\sum_{k=1}^{\infty} CB(k + d, \gamma) = \sum_{s=0}^{\infty} CB(s + 1 + d, \gamma) = 1)$. From that $\sum_{s=0}^{\infty} \frac{(1+d)_s}{(1+d+\gamma)_s} = 1/(CB(1 + r, \gamma))$. The LHS is Gauss Hypergeometric Function and Gauss hypergeometric theorem can be applied to get rid of the summation symbol. Simplifying and solving for C then gives the result.

	# patents	# categories	$\langle k \rangle$	Waring		Log-normal	
				γ	d	mean	std. dev.
1981	378,040	84,834	4.46	4.44	7.49	1.05	0.90
1986	702,508	99,776	7.04	4.00	11.27	1.40	1.02
1991	1,145,796	109,268	10.49	3.74	15.85	1.71	1.12
1996	1,654,321	114,372	14.46	3.43	18.89	1.96	1.20
2001	2,390,834	117,772	20.30	3.01	19.63	2.20	1.29
2006	3,209,056	119,833	26.78	2.69	18.96	2.36	1.36

Table 6.1: Parameter estimates for the size distribution at the level of subclasses.

6.4.2 Rank-size relationship of classes

The number of classes is quite low, which makes the probability density rather noisy.[5] For this reason, I choose to study the rank-size distribution. Carnabuci (2013) noticed that this relationship is not Zipfian. He fitted an exponential function, that is,

$$r(g) = Ce^{-xg},$$

where r is the rank and g is the relative size (i.e. the number of patents in the category of rank r, divided by the total number of patents[6]). However, while this fit is better than that of a power law and gives a good first approximation ($R^2 \approx 0.96$), the exponential functional form clearly underestimates the tail (as can be seen in the plots, but is only poorly reflected by the R^2). On the other hand, since the tail is not as fat as Zipf's law would imply, one needs to find an alternative functional form. Martínez-Mekler et al. (2009) showed that for a number of datasets the size-rank relation could be very well fitted by the following formula, which expresses size g as a function of rank r:

$$g(r) = K_1(N + 1 - r)^b r^{-a}, \tag{6.2}$$

where K_1 is a normalization constant. To ease later comparison with a close formula derived here, I consider $N + 1 \to N$. Then equation 6.2 can be rewritten as

$$g(r) = K_2 \left(1 - \frac{r}{N}\right)^b r^{-a}, \tag{6.3}$$

where K_2 is a normalization constant. This formula is like a power law, except for the factor $\left(1 - \frac{r}{N}\right)^b$ which decreases as the rank r approaches the number of categories N, creating an overall decay faster than normal power laws. Martínez-Mekler et al. (2009) call it a Generalized Beta Distribution (GBD).

---

[5]Sanditov (2005) fitted a log-normal and gamma distribution to the 1963-1999 database.

[6]Carnabuci (2013) worked with absolute size, but it is easier here to work with relative size.
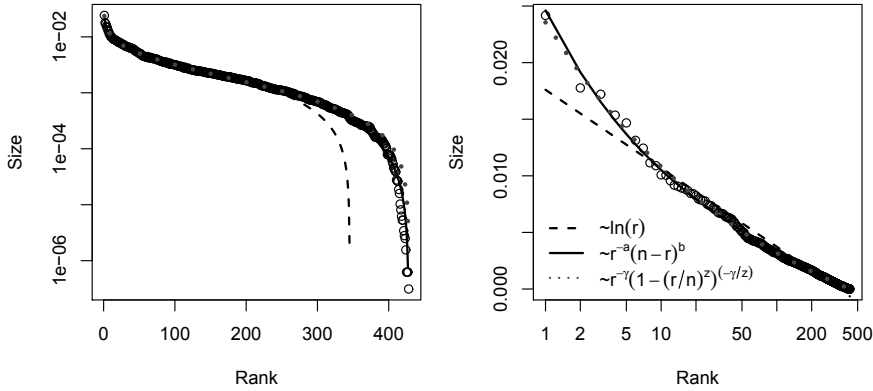
Figure 6.2: Fit of the size-rank relationship at the level of classes. The same data is displayed on the two panels. On the left panel, only the axis for size is on a log scale, as in Martínez-Mekler et al. (2009), and on the right, only the axis for rank is on a log scale, as in Carnabuci (2013).

	# of patents	# of categories	equation 6.3		equation 6.10	
			a	b	γ	z
1981	378040	423	0.33	1.14	0.24	0.28
1986	702508	425	0.35	1.08	0.27	0.34
1991	1145796	426	0.36	1.07	0.29	0.36
1996	1654321	427	0.35	1.13	0.27	0.33
2001	2390834	427	0.36	1.18	0.27	0.31
2006	3209056	428	0.36	1.36	0.23	0.23

Table 6.2: Parameter estimates for the size-rank relation at the level of classes.

Using the NBER patent data at the class level, I compared three hypothesis: the exponential relation between rank and size (Carnabuci 2013), the Generalized Beta Distribution (GBD) of Martínez-Mekler et al. (2009) (equation 6.3), and the slightly different GBD equation 6.10 derived in section 6.5 *infra*. Figure 6.2 shows that the two GBD clearly outperform the exponential. The R^2 values are 0.9597 for the exponential, 0.9974 for the GBD (6.3), and 0.9978 for the GBD derived from an original model in section 6.5 (equation 6.4). These results suggest that these two different GBD perform equally well, at least for this data. Note that following Martínez-Mekler et al. (2009), I take N directly from the data, that is, N is equal to the number of classes. Hence only two parameters are estimated by the fitting procedure.[7] Table 6.2 shows the estimated parameters for the two GBD over time.

---

[7]I used nonlinear least squares with a Gauss-Newton algorithm. Alternative optimization algorithms and starting conditions do not change the results significantly. Note that for Martínez-Mekler et al.'s (2009) equation, using (6.2) instead of (6.3) makes OLS estimation possible, by taking logs. In this case, the estimated parameters are slightly different (for 2006, $a = 0.24$ and

It is interesting to see that the parameters seem very stable for the first four 5-years periods, and change after 2001, except for the power law exponent (a) of Martínez-Mekler et al.'s (2009) equation, which is strikingly constant.[8] This may reflect a great stability in at least part of the process, as well as a slightly better performance of Martínez-Mekler et al.'s (2009) equation over the one presented here.

6.5 Theoretical models

Yule (1925) proposed the first model generating a power-law size distribution. He assumed that the number of categories grows at an exponential rate, and each category grows at an exponential rate.[9] Under these assumptions, the size distribution of categories is a Yule distribution, which has power law tails. Simon (1955) proposed a different version of Yule's model, in which time is not clock time but system time. However, up to a small modification, the two processes are equivalent (Simkin & Roychowdhury 2011). Naranan (1970) proposed a deterministic version and obtained similar results.

As we have seen in the previous section, neither the subclasses nor the classes have a pure power law size distribution. How should the Yule-Simon-Naranan's model be modified to produce the distributions observed empirically? To answer this question, I use Egghe's (2012) Generalized Naranan's Framework. Egghe (2012) extended Naranan's work for arbitrary invertible growth functions. He concluded his paper by asking whether one could obtain Martínez-Mekler et al.'s (2009) formula (equation 6.3) using the generalized Naranan's framework. Here I show first how to derive a *shifted* power law size distribution from the Generalized Naranan's framework (to explain the phenomenology of subclasses), and second how to derive a generalized beta curve very close to that of Martínez-Mekler et al. (2009) for the size-rank function (to explain the phenomenology of classes).

Consider that the number of categories at (continuous) time t is $\phi(t)$, and the number of items in a category of age t is $\psi(t)$. Moreover, assume that ϕ and ψ are invertible. Egghe (2012) proved the following:

Theorem 6.5.1. *In the generalized Naranan's framework, the size-rank function,*

$b = 1.55$, $R^2 = 0.986$).

[8]The same heuristic conclusion could be reached by looking at the plots of the four parameters over time for each year, not reported here.

[9]Yule's model is actually more complicated, in that he does not *assume* exponential growth but derives it from first principles, namely that each new species (item) has a fixed probability of generating a new species of a new genera and a fixed probability of generating a new species of an existing genera. Note that Yule's model is stochastic, but the deterministic version here can be seen as a mean-field approximation.

which gives the size of a category as a function of its rank is

$$G(r) = \psi(t - \phi^{-1}(r)).\tag{6.4}$$

Theorem 6.5.2. *In the generalized Naranan's framework, the frequency-size function, which gives the number of categories having a given size, is*

$$P(k) = \frac{\phi'(t - \psi^{-1}(k))}{\psi'(\psi^{-1}(k))}.\tag{6.5}$$

6.5.1   Shifted power law for subclasses

Let us start by analyzing the subclasses. We seek for ϕ and ψ such that $P(k) \propto (k + a)^{-\gamma}$. It is known from growing network models that a shifted attachment kernel (i.e. where the probability that the next newborn node links to an existing node of degree k is proportional to $k + constant$) gives a shifted power law. In other words, if a category accumulates items at a rate proportional to how many items it already has, growth is exponential and the size distribution is a power law. But if a category accumulates items at a rate proportional to how many items it already has *plus a constant*, we should expect a shifted power law. Hence, I assume

$$\frac{d\psi(t)}{dt} = x_1\psi(t) + x_2,$$

which implies

$$\psi(t) = \frac{-x_2 + e^{tx_1}(x_1 + x_2)}{x_1} = -q + c_2 e^{a_2 t},\tag{6.6}$$

where the parameters have been rewritten in a more condensed way. My assumption for the growth of the number of categories is the same as Yule-Naranan's:

$$\phi(t) = c_1 e^{a_1 t}.\tag{6.7}$$

Theorem 6.5 with the assumptions (6.6) and (6.7) gives

$$P(k) = \frac{a_1 c_2^{a_1/a_2} c_1 e^{a_1 t}}{a_2}(k + q)^{-1-\frac{a_1}{a_2}}.$$

The probability density $p(k)$ is obtained by diving by the total number of categories at time t,

$$p(k) = P(k)/\phi(t) = \frac{a_1}{a_2}(k + q)^{-1-\frac{a_1}{a_2}},$$

which is the shifted power law we were seeking to obtain. Moreover, it is a steady-state result. So, to obtain a shift in the power law size distribution in the generalized

Naranan's framework, all that needs to be modified is the shift parameter q in the function ψ describing the growth of individual categories. It implies that in their early life, categories (here, patent subclasses) benefit from an additional growth factor that ultimately dies out. This factor is equivalent to Dorogovtsev et al.'s (2000) "initial attractiveness", which is necessary when such models are applied to citation networks (de Solla Price 1976). In fact, using a stochastic and discrete model (that is, Simon's (1955) model with a shifted attachment kernel) would allow to derive directly the Waring distribution. However, I preferred to use the generalized Naranan's framework here, because it is much easier to apply to the next case of interest: the generalized beta size-rank function observed for patent classes.

6.5.2 Generalized beta for classes

To obtain a generalized beta size-rank function, one can stick to Naranan's assumption that individual categories grow exponentially, but needs to assume a very flexible function for the number of categories. More precisely, the number of items in a category of age t is

$$\psi(t) = c_2 e^{a_2 t}, \tag{6.8}$$

as in Naranan's, and the number of categories is Richard's curve. It is obtained as the solution of (Tsoularis & Wallace 2002)

$$\frac{d\phi(t)}{dt} = a_1 \phi(t) \left(1 - \left(\frac{\phi(t)}{n} \right)^z \right),$$

which is

$$\phi(t) = \frac{\phi_0 n}{\left(\phi_0^z + e^{-a_1 z t} \left(n^z - \phi_0^z \right) \right)^{\frac{1}{z}}}, \tag{6.9}$$

where the initial condition is $\phi(0) = \phi_0$. With $z = 1$, it is the classical logistic (Verhulst) equation, which is also S-shaped but symmetric. Verhulst equation is too simple to obtain a good rank-frequency function, because the inflexion point is precisely at the point where size has reached exactly half of the total possible (long term) size. The parameter z allows greater flexibility as to the position of the inflexion point, making the S-shaped asymmetric.

Theorem 6.5 under the assumptions (6.8) and (6.9) gives the following size-rank relation

$$G(r) = C_1 \left(1 - \left(\frac{r}{n} \right)^z \right)^{\gamma/z} r^{-\gamma},$$

where $C_1(t) = c_2 e^{a_2 t} (n\phi_0)^\gamma (n^z - \phi_0^z)^{-\gamma/z}$ and $\gamma = a_2/a_1$. Note that this equation will give 0 at the boundary $r = n$, which is not desirable but comes from the assumption that growth functions are continuous in this framework. The total size

(total number of patents) is $\sum_r G(r)$ (or the integral if we stick to the continuous approximation). Hence, since the relative size is $g(r) = G(r)/\sum_r G(r)$, we have

$$g(r) = C_2 \left(1 - \left(\frac{r}{n} \right)^z \right)^{\gamma/z} r^{-\gamma}, \tag{6.10}$$

where $C_2(t)$ is a normalization constant. Equation 6.10 is not exactly the same as, but is very similar to equation 6.3. Therefore, I believe that the mechanism described here provides a good theoretical background for the original equation proposed by Martínez-Mekler et al. (2009) (equation 6.2). Hence, the results presented here answer to a good extent the open problem stated in Egghe (2012), who asked if equation 6.2 could be derived from the Generalized Naranan's framework.

It should be noted that Martínez-Mekler et al. (2009) proposed a generative model for their equation, based on a dynamical model of boolean replication. Their model assumes that a vector of 0 and 1 is grown with a probabilistic rule, such that an entry is flipped with some probability, or repeated otherwise. They then show that the frequency with which non-overlapping sextuplets occur is well fitted by a GBD. Their interpretation is that the parameter a is associated to permanence (when an entry is replicated), whereas b is associated to change (when an entry if modified). The model presented here presents several differences with Martínez-Mekler et al.'s (2009). First, it is simpler. Second, the derivation of the GBD is analytical. Third, the formula is not exactly the same. Finally, the interpretation of the parameters is quite different. In the model of this paper, the parameters are determined directly by the parameters of the growth functions ϕ and ψ.

These results show that, using the generalized Naranan's framework, a non linear modification of the growth of the number of categories is enough to obtain a very good fit of the rank-size relationship in many cases. This does not mean that the number of categories is actually bounded, nor that we need to observe the growth functions assumed here to obtain the GBD. In fact, note that the variable t appears only in the constant of equation 6.10. Hence what the model means is that, whatever the growth functions are in clock time (say $\phi(\tau)$ and $\psi(\tau)$), if we can find a variable $t(\tau)$ such that $\phi(t)$ and $\psi(t)$ follow (6.8) and (6.9), then we should obtain the steady-state size-rank relation (6.10). It is difficult to test directly for the shape of ϕ and ψ here, because the data covers only about the last half of the US patents, while most classes were created very early, way before 1976 (Strumsky et al. 2012). Nevertheless, we can get some insights into the shape of ψ by taking logs of equation 6.8 to obtain the linear regressions:

$$\log \psi(t) = \log c_2 + a_2 t + \epsilon_t, \tag{6.11}$$

where t is the year. Figure 6.3 shows the results for the 415 classes which had positive size in 1976. We can conclude from it that exponential growth, while very rough and unfair to the heterogeneity among classes, is a decent assumption, with most R^2 above 0.9, and the coefficients highly concentrated between 0.05 and 0.15. Although this exercise does not show that growth is exponential, it suggests that the assumption of exponential growth has a fairly good amount of descriptive power. Finally, note that Richards' curve (6.9) cannot be meaningfully estimated directly, since in 1976 there were already 415 classes out of 428. The next section discusses why such simple assumptions nevertheless allow to make a good prediction of the size-rank relationship.
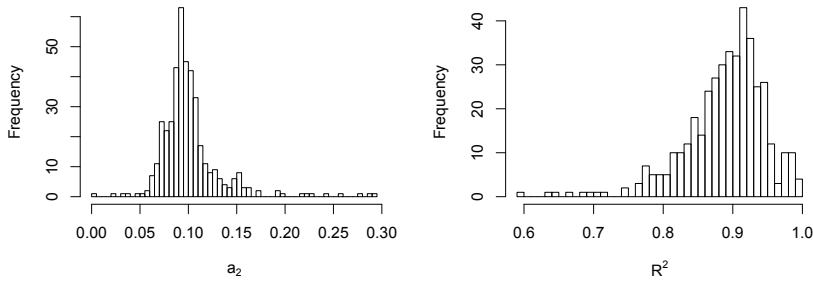

Figure 6.3: Test of the explanatory power of the exponential growth assumption of individual classes. OLS regressions for equation 6.11. The mean a_2 is 0.10 and the median 0.095.

6.6   Discussion

The assumptions of the models presented here should not be taken at face value. For instance, it is likely that the functions need not be deterministic for the results on the shape of the distributions to hold true (for the same reason that Yule and Simon's models give a result similar to Naranan's, which can be seen as deterministic mean-field continuous approximation of these discrete stochastic models). Moreover, what matters is not the exact shape of the growth function over (real) clock time, but the relationships that the time variable implies between the growth of different items and the number of items.[10] For these two reasons, one can expect that the models

---

[10]To be more explicit, consider a given model with a given $\phi(t)$ and $\psi(t)$, predicting $g(r)$ and $p(k)$. For the model to be correct, we need not observe ϕ and ψ as assumed, with t being our experienced time in months or years. What we need to observe is ϕ and ψ such that there exists a rescaling of time which makes them behaving as assumed over the rescaled time. It should be emphasized that, even for subclasses, the growth assumptions are not validated if we consider t

presented here can be stated in more general terms. Pinning down the general conditions under which these laws hold is certainly an interesting avenue for future research.

It can be seen in table 6.1 and 6.2 that the estimated exponents are not always constant over time. To the extent that this is true, the models used are falsified. The models display "steady-state" results, and therefore, if they were entirely correct, we would not observe time-varying parameters. However, one possibility is that the systems have not yet reached their steady-state, so that we are observing the convergence dynamics, which may not be accounted for in the models. Alternatively, and perhaps more plausibly, the system is actually in disequilibrium. There are no steady-state distribution of the US patent classes and subclasses sizes, because the way in which categories grow individually and in number is unstable. Note, in particular, that the mean number of patents per category is increasing. This trend can be worrying if one takes a realist view on technological categories. For instance, if one thinks that new categories reflect radical innovation and new patents in existing categories represent incremental innovations, then the data tell us that there is a tendency towards innovation being more and more incremental (less and less radical). echoing the concerns raised by other researchers using different theoretical frameworks (Jones 2009, Gordon 2012).

Finally, the models presented here shed light on the significant differences between the dynamics of classes and subclasses in the US patent classification system. Why is it the case that subclasses are almost power law distributed, whereas classes have a much less fat tail? The models suggest a simple explanation for that: the number of classes stops growing as the number of patents do, whereas the number of subclasses does keep a pace similar (in terms of functional form) to the growth of the number of patents. This is not a surprising result, because while classes have a purely horizontal structure (each class is at the same level of the classification tree), subclasses do not. Indeed, as explained in the literature review, subclasses have different indentation levels, so that as patents accumulate one can create a new subclass in a lower layer of the tree. Hence, if technological change is bounded horizontally (taking a realist view, there is a finite number of categories of natural phenomena, hence a finite number of categories of technologies based on them), this will be reflected in the number of classes but not in the number of subclasses. Moreover, while classes give an indication of the technological domain in which a patent is granted, subclasses are a mandatory classification for *claims*, which are the elements of novelty in a patent document. Novelty by definition cannot be novel for too long, hence as new patents arrive it should be expected that new subclasses

---

as years (e.g. the number of subclasses grows much slower than exponentially)

are created to accommodate genuine novelty. Beside these explanations, we cannot dismiss other factors behind the emergence of the US patent classification structure. It is possible that the decision to create or not a category is based on search efficiency arguments more than on ontological realities regarding the technology being described. For instance, it can be decided by classification decision-makers that the classes should not be changed too often, so as to provide a stability that is necessary to fix common understanding and expectations. This would explain the low growth of the number of classes. Albeit more speculative due to the lack of data, the non-fully-realist interpretation reminds us that patent classifications do not show directly the mesoscale organization of inventions, but the technologists' understanding of it, codified for specific purposes which are not only ontological but also functional (in particular searchability).

6.7 Conclusion

This chapter proposed three results. First, considering the US patents granted between 1976 and 2006, the size distribution of subclasses is well fitted by a shifted power law. Second, the size-rank relation of classes is well fitted by two different forms of a generalized beta distribution. Third, it is possible to derive a generalized beta functional form for rank-size relationships using the generalized Naranan's framework, by assuming an asymmetric logistic growth of the number of categories.

Much work is left to be done in this area. The models used make very strong assumptions but it remains to be seen how much conclusions would change if more randomness and heterogeneity were allowed. The role of time also needs to be clarified by mathematical arguments, both to understand how general are the models based on the generalized Naranan's framework, and to allow for disequilibrium dynamics, a common feature of knowledge systems. Moreover, the dataset used here covers (utility) patents granted in the US between 1976 and 2006, which is only about a half of the total number of US patents, whereas the models assume that the complete system is observed (notably the growth of the classification system). By observing these "universal" laws more precisely and in different contexts, future research may be able to understand better their fundamental drivers.

Chapter 7

Conclusion

At the scale of human evolution, the last two centuries of economic growth have been impressive (Maddison 2007). The exponential growth of the post WWII period generated hope that a sustained path of growth is achievable. Meanwhile, the alarming pressure on natural resources led to serious concerns about the possibility of exponential growth in a bounded world. But if infinite exponential growth of the extraction of natural resources is impossible, what can we hope to keep growing forever?

The value of the economic approach is to consider welfare, without necessarily equating welfare with an increased consumption of natural resources intensive goods. As many have argued (e.g. the conclusion in Arthur (2009)), technologies create new constraints but on average they tend to increase human capabilities. Interestingly, the literature on alternative indicators of welfare tends to consider knowledge in itself as valuable for development (Sen 1999). The primacy of knowledge as a source of economic progress is well reflected in the debates about the actual possibility of long run growth, which often boil down to an opposition between technology optimists and pessimists. This debate has been particularly alive in the last years, following Gordon's (2012) claim that innovation is slowing down. But for technology optimists, such as Brynjolfsson & McAfee (2014), "secular stagnation" is unlikely in view of the potentially tremendous transformative impact of new technologies. This debate is of course not new; in one form or another it has been going on since Malthus. What seems to generate consensus, however, is that the answer lies in our ability or inability to produce, allocate and use knowledge (Soete 2009, Fernald & Jones 2014).

This thesis took as a starting point the idea that development is knowledge-based. It argued that as a result, the focus of economics should shift from the self-organization of markets to the self-organization of knowledge systems. Therefore, it proposed to study in detail the structure and evolution of abstract knowledge

economies – agents and their ideas – but also more concrete examples, such as scientific citations or patent classifications . Knowledge systems should be studied in their own right, because they are very *complex*, and quite different from traditional economic systems (textbook's production and exchange economies). Even more than "normal" goods and services, knowledge or "ideas" evolve in a way that involves feedback loops, self-reinforcing processes, auto-catalytic dynamics, and leads to structures that have non trivial properties. These properties – power law distributions have been the prime example in this work – can be nevertheless well-understood once we recognize them as signatures of complex systems.

A particularly interesting connection between power laws and knowledge systems is that power laws are known to emerge when a system features both a growing number of elements and growth of these elements. Hence, the nature of knowledge – which is non-rival, builds out of itself, and is sometimes radically (essentially) new – makes stochastic models for power law particularly attractive from a theoretical point of view. Structural change, in this framework, can be seen as the creation, diffusion and selection of varieties, be they product varieties, capital vintages and blueprints, or behavioral and cognitive rules. Necessarily, to be empirically relevant, classical models such as Simon (1955) are often too simple and need to be improved. In the present work, I have focused on some nonlinearities, emerging either from theoretical considerations (as bounded population in chapter 4) or empirical observations (as the non power law scaling of the size distribution patent classes in chapter 6). These are only a few instances of the ways in which stochastic models in this tradition can provide both theoretical and empirical insights into the process of development. Other prominent examples include of course size distributions (such as that of firms, sectors, and cities), but also the degree distributions in consumer-, firm-, and country-product networks. The latter often exhibit a "nested" structure which reflects the pattern of growth implied by the innovation/diffusion duality.

To summarize the thesis more specifically, I have proposed parsimonious theoretical and empirical models of self-organizing knowledge systems. Chapter 4 characterized precisely the self-organization of innovation - diffusion economies. Chapter 5 derived predictions of this model on empirical observables, arguing that the main driver of the structure of citation networks is the way in which and frequency with which knowledge diffuses. Chapter 6 studied the size of technological categories, showing that an S-shaped growth of the number of categories is a sufficient modification of existing models to explain the rank-size relationship at the level of patent classes.

There are two general lessons from this thesis, one technical and one theoretical. From a technical point of view, it has been argued that the theory of self-organizing

two-mode networks provides a novel and useful framework to study economic processes. In particular, it was shown that one can study the co-evolution of social and epistemic networks by keeping track of the two-mode agent-idea network. This framework is rather general, and the original models presented here are only early examples of a mathematical theory which will be developed in the years to come, and for which the economic theory is largely to be written.

The theoretical lesson from this thesis concerns the non-equilibrium nature of knowledge systems, which casts a doubt on their viability. As knowledge accumulates, can it keep growing at the same pace? Chapter 5 successfully fitted a distribution derived from an "equilibrium" model, but could not show that the parameters were constant over time. Equilibrium in the models that I presented is a "statistical equilibrium", that is a state in which many things keep happening (contrary to Nash or Walras equilibria), but compensate in such a way that mesolevel quantities (such as certain distributions) are stable. However, the change of these fits over time indicates that citation networks cannot be fully characterized by (even statistical) equilibrium models. In chapter 6, a strong nonlinearity in the growth of the number of categories was emphasized. In fact, during the 30 years studied, only a few classes were added in spite of a tremendous increase in the number of granted patents. In both cases, these nonlinearities and disequilibria suggest that the structure of knowledge systems changes as knowledge systems grow, which opens the question of their viability. As a result, understanding the sources and consequences of these disequilibria and nonlinearities is an important challenge for future research.

Chapter 8

Societal Relevance

In accordance with article 23.5 of the "Regulation governing the attainment of doctoral degrees at Maastricht University" decreed by resolution of the Board of Deans, dated 3 July 2013, an addendum must be added about valorization.

At any level, from pure theory to public policy and private practice, it is thought that knowledge is at the heart of economic progress. Knowledge allows persistent productivity increases which ensure us access to more goods, services, and free time. These ideas often translate into formal quantitative objectives. a famous example being the Lisbon Agenda in 2000, which aimed at making the E.U. "the most dynamic and competitive knowledge-based economy in the world by 2010", by requiring that public and private Research and Development (R&D) spending should attain a level 3% of GDP.

If, really, what matters is the accumulation of knowledge, then one must study in detail the microeconomic dynamics of knowledge creation and diffusion. Of course, understanding the organization of markets and the relations among aggregated variables is relevant and useful. But once one admits the primacy of knowledge, innovation, technological and institutional change in explaining the long-run path of the human race, it seems imperative to try understanding the self-organization of knowledge systems. How is knowledge created? How does it diffuse? How does knowledge creation fuel additional creation? How does knowledge diffusion influence the dynamics of knowledge accumulation? How do different pieces of knowledge relate to each other?

If we understand what are the key parameters driving the evolution of knowledge systems, we may be able to measure, predict, and control them. The relevance of this thesis is right here: identifying the most fundamental parameters and principles of organization and measuring them on real systems to understand the real-world dynamics. The controllability, or normative analysis of knowledge systems has

been only superficially touched upon in the present work. Instead, I have focused on giving basic theoretical foundations and data-constrained empirical evidence for the evolution of knowledge systems. Let me now highlight what types of policy-oriented perspectives follow from this work.

First of all, understanding the structure of knowledge based on its dynamics allows potentially to predict which type of knowledge will emerge. To a very large extent, knowledge is created from existing knowledge, which implies that we can think of a structure, or system, or network of ideas. In the last decades, it has become clear that in some specific cases, we can actually describe and understand the structure of knowledge using the powerful mathematical concept of graph (or network). This is possible because codified knowledge items often formally refer to other pieces of knowledge that have been influential or are epistemically related. Materially speaking, pieces of codified knowledge such as scientific papers and patents "cite" other documents, such that the actual structure of knowledge can be described as a citation network, where each knowledge item is a node and each node has a directed link towards each of the knowledge items that it cites. As it turns out, the observed empirical patterns are striking and the associated theory is insightful: most citation networks have a power law distribution of in-degree, which can be explained by growth and cumulative advantage (or "rich gets richer" effects). In more mundane terms, the vast majority of knowledge items are almost never cited, whereas a few knowledge items are referenced a very high number of times (a popular rule of thumb for the so-called "Pareto principle" would phrase it as "20% of the papers gather 80% of citations", though the actual numbers are slightly different). This observation can be explained by the fact that the number of ideas (say scientific papers or patents) is growing, and there exists a cumulative advantage, or rich gets richer, or Matthew effect at play: well recognized ideas tend to diffuse even more, creating a snowballing effect or self-reinforcing process that eventually leads to this high level of inequality among knowledge items in terms of diffusion or being "cited".

Understanding such processes will allow to make better predictions. For instance, the study of citation networks can result in predictions of the future success of particular items, or prediction of which knowledge domains will grow more. A sound understanding of cause and effect may even permit to detect inefficiencies and propose policies to correct them. Relatedly, it is important to appraise how and why scientific and technological credit are so unevenly distributed, because they are often used for the evaluation of individuals, laboratories, firms or projects. In this thesis, it is shown that imitating a random idea of a random friend (independent of the idea's quality) leads to a structure of the citation network fairly close

to the observed one. Of course, in reality some papers are better than others and they receive more citations *because* they are better; but as a matter of fact, to a first approximation we don't need to invoke the quality argument to understand why citation distributions are so unequal. Hence, this work warns evaluators that using citations as an indicator of scientific quality can be misleading, emphasizing that other factors, such as the social structure in which diffusion takes place, or the allocation of attention between innovation and imitation, represents a plausible explanation for most of the observed reality.

Another focus of this thesis has been classification systems. The analysis of classification systems has a tremendous economic and social relevance because categorizations are pervasive, and wherever they are used, they constrain their users and influence the future dynamics of the system being categorized. In the present work, I analyzed the classification of US patents, arguing that the addition of new categories reflects the fact that some patents are *sui generis* (of their own kind), and as such can be considered very innovative. The dynamics of the category system then informs us about the evolution of the degree of novelty of inventions. Of course, this reasoning reflects the belief that patent officers are reliable at classifying patents and creating new categories. But, presumably, other patent officers may have made a different classification choice; different countries with different cultures and times of entry in the industrial revolution would develop different classification systems. Along this line of reasoning emerges the question of standardization of classification systems or benefits of their multiplicity. By showing that patent classification systems can be used to ask macro- (or meso-)economic questions about the dynamics of knowledge systems and their future, this work suggests that the benefits of non-standardisation of classification may have been underestimated when it was decided to merge the different patent classification systems (International, US and EU) – currently an ongoing process. While the purpose of this work was not to decide whether or not the risk of cognitive lock-in is higher than the benefits of standardization, it contributes to raising the question. This thesis uncovered how the size of patent classes (in terms of the number of patents) is distributed, and had to propose an original, but simple and not fully explanatory model for it. This only shows our relative ignorance of classification dynamics, and the large amount of work which remains to be done before decision on standardization be based on scientific arguments. Moreover, relatively recent technological trends are transforming the way objects are classified. Information technologies, coupled with user-driven (so called Web 2.0) dynamics, potentially change radically the way things (material or not) are classified, because electronic classification facilitates the assignment of multiple categories, and because users tagging objects effectively

define categories in a decentralized way. Theses changes have potentially important effects on society and the economy – yet if little is known on statistical properties of existing classification systems, understanding these changes will prove difficult. In this regard, the analysis of patent categories in chapter 6, and the technicalities of chapter 3 and 4 (which may be used to model folksonomies) may contribute to a better understanding and design of information systems.

Another practical domain for which this thesis is relevant concerns recommender systems. If we start from the idea that attention is limited, and that what goes to people's attention is important, we have to understand the socio-economic consequences of different types of recommender systems. To obtain the normative results that would be useful for policy makers, we need to understand exactly how the structure of user-objects systems (bipartite or two-mode networks) is influenced by different types of recommendations (e.g, recommending the most popular item, recommending items already used by friends, recommending completely at random, etc.). Chapter 3 and 4 contributed to the technical literature in this area, which was so far limited to the physics and computer science communities. Hopefully, raising the interest of economists will ultimately lead to more rigorous positive and normative analysis of the individual and collective welfare properties of different recommendation algorithms, leading to potentially huge long-term welfare gains.

Finally, there is one mathematical truth, an identity, in the footnote of page 62, that has been discovered. Ironically, its discovery is fully serendipitous, and has had almost no consequence for the scientific topics of this thesis. In principle, it can be embodied in commercial softwares. This equality is very specific, and in all likelihood it will not be useful to many people. Perhaps it is not even worth the cost of the programmer who would have to implement it. Nevertheless, if it was implemented and if people would use it, (human and machine) computing time and energy would be saved. This point demonstrates that serendipity exists, and serves to express the view that the social relevance of fundamental sciences should be evaluated taking into account that real truth have to be proven only once. As such they have a great socio-economic value, possibly more than their marginal cost –but more theoretical work on the microeconomics of knowledge would be necessary to establish or falsify this statement.

Bibliography

Abramovitz, M. (1956), 'Resource and output trends in the United States since 1870', *The American Economic Review* **46**(2), pp. 5–23.

Adamic, L. & Huberman, B. (2000), 'Power-law distribution of the world wide web', *Science* **287**(5461), 2115–2115.

Aghion, P. & Howitt, P. (1992), 'A model of growth through creative destruction', *Econometrica* **60**(2), pp. 323–351.

Albarrán, P. & Ruiz-Castillo, J. (2011), 'References made and citations received by scientific articles', *Journal of the American Society for Information Science and Technology* **62**(1), 40–49.

Alchian, A. A. (1950), 'Uncertainty, evolution, and economic theory', *Journal of Political Economy* **58**(3), 211–221.

Andersen, B. (1999), 'The hunt for S-shaped growth paths in technological innovation: a patent study', *Journal of Evolutionary Economics* **9**(4), 487–526.

Ando, A., Fisher, F. M. & Simon, H. A. (1963), *Essays on the structure of social science models*, MIT Press, Cambridge, MA.

Armenter, R. & Koren, M. (2014), 'A balls-and-bins model of trade', *American Economic Review* **104**(7), 2127–2151.

Arrow, K. (1962), Economic welfare and the allocation of resources for invention, *in* 'The rate and direction of inventive activity: Economic and social factors', NBER, pp. 609–626.

Arrow, K. J. & Debreu, G. (1954), 'Existence of an equilibrium for a competitive economy', *Econometrica* **22**(3), 265–290.

Arthur, B. W., Ermoliev, Y. M. & Kaniovski, Y. M. (1987), 'Path-dependent processes and the emergence of macro-structure', *European journal of operational research* **30**(3), 294–303.

Arthur, W. (1989), 'Competing technologies, increasing returns, and lock-in by historical events', *The Economic Journal* **99**(394), 116–131.

Arthur, W. (2009), *The Nature of Technology: What It Is and How It Evolves*, Free Press.

Arthur, W. B. (1994), *Increasing returns and path dependence in the economy*, University of Michigan Press.

Atalay, E. (2013), 'Sources of variation in social networks', *Games and Economic Behavior* **79**, 106–131.

Atkin, R. (1977), *Combinatorial connectivities in social systems*, Birkhauser Verlag.

Axelrod, R. M. (1997), *The complexity of cooperation: Agent-based models of competition and collaboration*, Princeton University Press.

Axtell, R. L. (2001), 'Zipf distribution of US firm sizes', *Science* **293**(5536), 1818–1820.

Bala, V. & Goyal, S. (1998), 'Learning from neighbors', *Review of Economic Studies* **65**, 595–621.

Bala, V. & Goyal, S. (2000), 'A noncooperative model of network formation', *Econometrica* **68**(5), 1181–1229.

Baldwin, C. & Clark, K. (2000), *Design rules: The power of modularity*, The MIT Press.

Barabási, A. & Albert, R. (1999), 'Emergence of scaling in random networks', *Science* **286**(5439), 509.

Becker, M. C., Cohendet, P. & Llerena, P. (2007), Division of labor and division of knowledge: why the nature of the causality matters for the evolutionary theory of the firm, *in* 'Innovation, Industrial Dynamics and Structural Transformation', Springer, pp. 49–63.

Beguerisse-Diaz, M., Porter, M. A. & Onnela, J.-P. (2010), 'Competition for popularity in bipartite networks', *Chaos: An Interdisciplinary Journal of Nonlinear Science* **20**(4).

Berge, C. (1989), *Hypergraphs: combinatorics of finite sets*, North-Holland.

Bianconi, G. & Barabási, A. (2001), 'Competition and multiscaling in evolving networks', *EPL (Europhysics Letters)* **54**, 436.

Biglu, M. (2008), 'The influence of references per paper in the SCI to Impact Factors and the Matthew Effect', *Scientometrics* **74**(3), 453–470.

Bollobás, B. (1998), *Modern graph theory*, Vol. 184, Springer.

Bollobás, B., Riordan, O., Spencer, J., Tusnády, G. et al. (2001), 'The degree sequence of a scale-free random graph process', *Random Structures & Algorithms* **18**(3), 279–290.

Bonacich, P. (1987), 'Power and centrality: A family of measures', *American Journal of Sociology* **92**(5), 1170–1182.

Börner, K., Maru, J. & Goldstone, R. (2004), 'The simultaneous evolution of author and paper networks', *Proceedings of the National Academy of Sciences of the United States of America* **101**(Suppl 1), 5266–5273.

Boschma, R. & Frenken, K. (2010), The spatial evolution of innovation networks. a proximity perspective, *in* 'Handbook of evolutionary economic geography', Edward Elgar Cheltenham, pp. 120–135.

Bottazzi, G., Dosi, G., Fagiolo, G. & Secchi, A. (2007), 'Modeling industrial evolution in geographical space', *Journal of Economic Geography* **7**(5), 651–672.

Bottazzi, G. & Secchi, A. (2006), 'Gibrat's law and diversification', *Industrial and Corporate Change* **15**(5), 847–875.

Bowker, G. & Star, S. (2000), *Sorting things out: classification and its consequences*, The MIT Press.

Bramoullé, Y., Currarini, S., Jackson, M. O., Pin, P. & Rogers, B. W. (2012), 'Homophily and long-run integration in social networks', *Journal of Economic Theory* **147**(5), 1754–1786.

Breiger, R. (1974), 'The duality of persons and groups', *Social Forces* **53**, 181.

Breschi, S. & Lissoni, F. (2009), 'Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows', *Journal of Economic Geography* **9**(4), 439–468.

Brynjolfsson, E. & McAfee, A. (2014), *The second machine age: work, progress, and prosperity in a time of brilliant technologies*, WW Norton & Company.

Burt, R. S. (1995), *Structural Holes: The Social Structure of Competition*, Harvard University Press.

Carayol, N. & Dalle, J.-M. (2007), 'Sequential problem choice and the reward system in Open Science', *Structural Change and Economic Dynamics* **18**(2), 167–191.

Carnabuci, G. (2013), 'The distribution of technological progress', *Empirical Economics* **44**(3), 1143–1154.

Castells, M. (2011), *The rise of the network society: The information age: Economy, society, and culture*, John Wiley & Sons.

Champernowne, D. G. (1953), 'A model of income distribution', *The Economic Journal* **63**(250), 318–351.

Chaney, T. (2011), 'The Network Structure of International Trade', *NBER working paper* **16753**(forthcoming).

Chung, F., Handjani, S. & Jungreis, D. (2003), 'Generalizations of Polya's urn problem', *Annals of combinatorics* **7**(2), 141–153.

Clauset, A., Shalizi, C. & Newman, M. (2009), 'Power-law distributions in empirical data', *SIAM Review* **51**(4), 661–703.

Cohen, W. & Levinthal, D. (1989), 'Innovation and learning: the two faces of R & D', *The Economic Journal* **99**(397), 569–596.

Collins, H. M. (1974), 'The tea set: Tacit knowledge and scientific networks', *Science studies* **4**(2), 165–185.

Cooke, P. (2001), 'Regional innovation systems, clusters, and the knowledge economy', *Industrial and corporate change* **10**(4), 945–974.

Cooper, C. & Prałat, P. (2010), 'Scale-free graphs of increasing degree', *Random Structures & Algorithms* **38**(4), 396–421.

Cowan, R. (1991), 'Tortoises and hares: choice among technologies of unknown merit', *The Economic Journal* **101**(407), 801–814.

Cowan, R. (2005), Network models of innovation and knowledge diffusion, *in* S. Breschi & F. Malerba, eds, 'Clusters, networks and innovation', Oxford University Press New York, pp. 29–53.

Cowan, R., David, P. & Foray, D. (2000), 'The explicit economics of knowledge codification and tacitness', *Industrial and Corporate Change* **9**(2), 211–253.

Cowan, R. & Jonard, N. (2004), 'Network structure and the diffusion of knowledge', *Journal of Economic Dynamics and Control* **28**(8), 1557–1575.

Cowan, R. & Jonard, N. (2009), 'Knowledge portfolios and the organization of innovation networks', *The Academy of Management Review* **34**(2), 320–342.

Criado, R., Flores, J., García del Amo, A., Gómez-Gardeñes, J. & Romance, M. (2012), 'A mathematical model for networks with structures in the mesoscale', *International Journal of Computer Mathematics* **89**(3), 291–309.

Csárdi, G., Strandburg, K., Zalányi, L., Tobochnik, J. & Érdi, P. (2007), 'Modeling innovation by a kinetic description of the patent citation system', *Physica A: Statistical Mechanics and its Applications* **374**(2), 783–793.

David, P. A. (1985), 'Clio and the economics of QWERTY', *The American Economic Review* **75**(2), 332–337.

David, P. A. & Greenstein, S. (1990), 'The economics of compatibility standards: An introduction to recent research', *Economics of Innovation and New Technology* **1**(1-2), 3–41.

de Sola Pool, I. & Kochen, M. (1979), 'Contacts and influence', *Social networks* **1**(1), 5–51.

de Solla Price, D. (1965), 'Networks of scientific papers', *Science* **149**(3683), 510.

de Solla Price, D. (1976), 'A general theory of bibliometric and other cumulative advantage processes', *Journal of the American Society for Information Science* **27**(5), 292–306.

Dopfer, K., Foster, J. & Potts, J. (2004), 'Micro-meso-macro', *Journal of Evolutionary Economics* **14**(3), 263–279.

Dorogovtsev, S. & Mendes, J. (2000), 'Evolution of networks with aging of sites', *Physical Review E* **62**(2), 1842.

Dorogovtsev, S. & Mendes, J. (2001), 'Effect of the accelerating growth of communications networks on their structure', *Physical Review E* **63**(2), 025101.

Dorogovtsev, S., Mendes, J. & Samukhin, A. (2000), 'Structure of growing networks with preferential linking', *Physical Review Letters* **85**(21), 4633–4636.

Dosi, G. (1982), 'Technological paradigms and technological trajectories:: A suggested interpretation of the determinants and directions of technical change', *Research Policy* **11**(3), 147–162.

Dosi, G., Marengo, L. & Fagiolo, G. (2006), Learning in evolutionary environments, *in* K. Dopfer, ed., 'The evolutionary foundations of economics', Cambridge University Pres.

Easley, D. & Kleinberg, J. (2010), *Networks, crowds, and markets*, Cambridge Univ Press.

Economides, N. (1996), 'The economics of networks', *International Journal of Industrial Organization* **14**(6), 673–699.

Egghe, L. (2012), 'Study of rank-and size-frequency functions and their relations in a generalized naranan framework', *Mathematical and Computer Modelling* **55**(7), 1898–1903.

Ehrhardt, G., Marsili, M. & Vega-Redondo, F. (2006), 'Diffusion and growth in an evolving network', *International Journal of Game Theory* **34**(3), 383–397.

Eom, Y. & Fortunato, S. (2011), 'Characterizing and modeling citation dynamics', *PLoS ONE* **6**(9), e24926.

Epstein, J. M. & Axtell, R. L. (1996), *Growing artificial societies: Social science from the bottom up*, The Brookings Institution Press.

Erdős, P. & Rényi, A. (1960), 'On the evolution of random graphs', *Publ. Math. Inst. Hung. Acad. Sci* **5**, 17–61.

Evans, T. (2007), 'Exact solutions for network rewiring models', *The European Physical Journal B* **56**(1), 65–69.

Evans, T. & Plato, A. (2008), 'Network rewiring models', *Networks and Heterogenous Media* **3**(2), 221–238.

Falasco, L. (2002), 'Bases of the United States Patent Classification', *World Patent Information* **24**(1), 31–33.

Faust, K. (1997), 'Centrality in affiliation networks', *Social Networks* **19**(2), 157–191.

Fenner, T., Levene, M. & Loizou, G. (2005), 'A stochastic evolutionary model exhibiting power-law behaviour with an exponential cutoff', *Physica A: Statistical Mechanics and its Applications* **355**(2), 641–656.

Fernald, J. G. & Jones, C. I. (2014), The future of us economic growth, Technical report, National Bureau of Economic Research.

Foucault, M. (1966), *Les mots et les choses*, Vol. 42, Gallimard Paris.

Freeman, C. (1987), *Technology policy and economic performance: lessons from Japan*, Pinter Publishers London.

Freeman, C. & Soete, L. (1997), *The economics of industrial innovation*, third edn, MIT Press, Cambridge, MA.

Freeman, L. (2003), Finding social groups: A meta-analysis of the southern women data, *in* 'Dynamic Social Network Modeling and Analysis: workshop summary and papers', pp. 39–77.

Frenken, K. & Boschma, R. A. (2007), 'A theoretical framework for evolutionary economic geography: industrial dynamics and urban growth as a branching process', *Journal of Economic Geography* **7**(5), 635–649.

Frenken, K., Marengo, L. & Valente, M. (1999), Interdependencies, nearly-decomposability and adaptation, *in* 'Computational techniques for modelling learning in economics', Springer, pp. 145–165.

Fu, D., Pammolli, F., Buldyrev, S. V., Riccaboni, M., Matia, K., Yamasaki, K. & Stanley, H. E. (2005), 'The growth of business firms: Theoretical framework and empirical evidence', *Proceedings of the National Academy of Sciences of the United States of America* **102**(52), 18801–18806.

Gabaix, X. (1999), 'Zipf's law for cities: an explanation', *The Quarterly Journal of Economics* **114**(3), 739–767.

Gabaix, X. (2009), 'Power laws in economics and finance', *Annual Review of Economics* **1**(1), 255–294.

Garibaldi, U. & Scalas, E. (2010), *Finitary probabilistic methods in econophysics*, Cambridge University Press Cambridge.

Gehrke, J., Ginsparg, P. & Kleinberg, J. (2003), 'Overview of the 2003 KDD cup', *ACM SIGKDD Explorations Newsletter* **5**(2), 149–151.

Geroski, P. A. (2000), 'Models of technology diffusion', *Research policy* **29**(4), 603–625.

Ghiglino, C. & Kuschy, N. (2011), 'Are patent citations driven by quality?', *Mimeo, University of Essex* .

Gibrat, R. (1931), *Les inégalités économiques*, Recueil Sirey.

Golosovsky, M. & Solomon, S. (2012), 'Runaway events dominate the heavy tail of citation distributions', *The European Physical Journal-Special Topics* **205**(1), 303–311.

Golub, B. & Jackson, M. O. (2010), 'Naive learning in social networks and the wisdom of crowds', *American Economic Journal: Microeconomics* **2**(1), 112–149.

Gordon, R. J. (2012), 'Is US economic growth over? faltering innovation confronts the six headwinds', *NBER Working Paper* **18315**.

Gosper, R. W. (1978), 'Decision procedure for indefinite hypergeometric summation', *Proceedings of the National Academy of Sciences of the United States of America* **75**(1), 40–42.

Goyal, S. (2007), *Connections: an introduction to the economics of networks*, Princeton University Press.

Granell, C., Gómez, S. & Arenas, A. (2011), 'Mesoscopic analysis of networks: Applications to exploratory analysis and data clustering', *Chaos: An Interdisciplinary Journal of Nonlinear Science* **21**(1), 016102.

Granovetter, M. (1973), 'The Strength of Weak Ties', *American Journal of Sociology* **78**(6), 1360–1380.

Grossman, G. M. & Helpman, E. (1991), 'Quality ladders in the theory of growth', *The Review of Economic Studies* **58**(1), 43–61.

Gutiérrez Jáimez, R. & Rodríguez Avi, J. (1997), 'Family of Pearson discrete distributions generated by the univariate hypergeometric function $_3F_2$ $(\alpha_1, \alpha_2, \alpha_3; \gamma_1, \gamma_2; \lambda)$', *Applied stochastic models and data analysis* **13**(2), 115–125.

Hall, B. H. & Rosenberg, N. (2010), *Handbook of the Economics of Innovation*, Elsevier.

Hall, B., Jaffe, A. & Trajtenberg, M. (2001), 'The NBER patent citation data file: Lessons, insights and methodological tools'.

Hawkins, D. & Simon, H. A. (1949), 'Note: some conditions of macroeconomic stability', *Econometrica* **17**(3/4), 245–248.

Hébert-Dufresne, L., Allard, A., Marceau, V., Noël, P.-A. & Dubé, L. J. (2011), 'Structural preferential attachment: Network organization beyond the link', *Physical Review Letters* **107**(15), 158702.

Hicks, D. (2011), 'Structural change and industrial classification', *Structural Change and Economic Dynamics* **22**(2), 93–105.

Hidalgo, C., Klinger, B., Barabási, A. & Hausmann, R. (2007), 'The product space conditions the development of nations', *Science* **317**(5837), 482–487.

Hirschman, A. (1958), *The strategy of economic development*, Yale University Press, New Haven, CT.

Hodgson, G. M. (2006), *The Evolutionary Foundations of Economics*, Cambridge University Press, chapter Decomposition and growth: biological metaphors in economics from the 1880s to the 1980s.

Hou, Z., Kong, X., Shi, D., Chen, G. & Zhao, Q. (2009), Degree-distribution stability of growing networks, *in* 'Complex Sciences', Springer, pp. 1827–1837.

Iguchi, K. & Yamada, H. S. (2007), 'General connectivity distribution functions for growing networks with preferential attachment of fractional power', *Journal of Mathematical Physics* **48**, 113303.

Ijiri, Y. & Simon, H. A. (1977), *Skew distributions and the sizes of business firms*, North-Holland Publishing Company New York.

Irwin, J. (1963), 'The place of mathematics in medical and biological statistics', *Journal of the Royal Statistical Society. Series A (General)* pp. 1–45.

Jackson, M. O. (2008), *Social and Economic Networks*, Princeton University Press.

Jackson, M. O. & Wolinsky, A. (1996), 'A strategic model of social and economic networks', *Journal of Economic Theory* **71**(1), 44–74.

Jackson, M. & Rogers, B. (2007), 'Meeting strangers and friends of friends: How random are social networks?', *The American Economic Review* **97**(3), 890–915.

Jaffe, A. B., Trajtenberg, M. & Henderson, R. (1993), 'Geographic localization of knowledge spillovers as evidenced by patent citations', *the Quarterly Journal of Economics* **108**(3), 577–598.

Johnson, N. L., Kemp, A. W. & Kotz, S. (2005), *Univariate discrete distributions*, Vol. 444, John Wiley & Sons.

Jones, B. (2009), 'The burden of knowledge and the death of the renaissance man: Is innovation getting harder?', *Review of Economic Studies* **76**(1), 283–317.

Jones, C. I. (1995), 'R&D-based models of economic growth', *Journal of Political Economy* **103**(4), 759–784.

Jones, C. I. (2005), 'The shape of production functions and the direction of technical change', *The Quarterly Journal of Economics* **120**(2), 517–549.

Jovanovic, B. & Rob, R. (1989), 'The growth and diffusion of knowledge', *The Review of Economic Studies* **56**(4), 569–582.

Kauffman, S. (1993), *The origins of order: Self organization and selection in evolution*, Oxford University Press.

Kirman, A. (1993), 'Ants, rationality, and recruitment', *The Quarterly Journal of Economics* **108**(1), 137–156.

Kivelä, M., Arenas, A., Barthelemy, M., Gleeson, J. P., Moreno, Y. & Porter, M. A. (2013), 'Multilayer networks', *arXiv preprint arXiv:1309.7233* .

Kleinberg, J. M., Kumar, R., Raghavan, P., Rajagopalan, S. & Tomkins, A. S. (1999), The Web as a Graph: Measurements, Models, and Methods, *in* T. Asano, H. Imai, D. T. Lee, S.-i. Nakano & T. Tokuyama, eds, 'Computing and Combinatorics', Vol. 1627 of *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, Berlin, Heidelberg, chapter 1, pp. 1–17.

Klepper, S. (1997), 'Industry life cycles', *Industrial and Corporate Change* **6**(1), 145–182.

Kline, S. J. & Rosenberg, N. (1986), An overview of innovation, *in* R. Landau & N. Rosenberg, eds, 'The positive sum strategy: Harnessing technology for economic growth', National Academy Press Washington, DC.

König, M., Lorenz, J. & Zilibotti, F. (2012), 'Innovation vs imitation and the evolution of productivity distributions', *CEPR Discussion Paper* **8843**.

Krapivsky, P. & Krioukov, D. (2008), 'Scale-free networks as preasymptotic regimes of superlinear preferential attachment', *Physical Review E* **78**(2), 026114.

Krapivsky, P., Redner, S. & Leyvraz, F. (2000), 'Connectivity of growing random networks', *Physical Review Letters* **85**(21), 4629–4632.

Krugman, P. (1996), *The self-organizing economy, Mitsui Lectures in Economics*, Wiley-Blackwell, Cambridge, Mass., USA.

Kuhn, T. (1962), *The structure of scientific revolutions*, University of Chicago Press.

Kullmann, L. & Kertész, J. (2001), 'Preferential growth: Exact solution of the time-dependent distributions', *Physical Review E* **63**(5), 051112.

Laherrere, J. & Sornette, D. (1998), 'Stretched exponential distributions in nature and economy:fat tails with characteristic scales', *The European Physical Journal B-Condensed Matter and Complex Systems* **2**(4), 525–539.

Latapy, M., Magnien, C. & Vecchio, N. (2008), 'Basic notions for the analysis of large two-mode networks', *Social Networks* **30**(1), 31–48.

Latour, B. (2005), *Reassembling the social: An introduction to actor-network-theory*, Oxford University Press, USA.

Latour, B. & Woolgar, S. (1979), *Laboratory life: The construction of scientific facts*, Princeton University Press.

Leskovec, J., Kleinberg, J. & Faloutsos, C. (2005), Graphs over time: densification laws, shrinking diameters and possible explanations, *in* 'Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining', ACM, pp. 177–187.

Levene, M., Fenner, T., Loizou, G. & Wheeldon, R. (2002), 'A stochastic model for the evolution of the web', *Computer Networks* **39**(3), 277–287.

Liu, C., Yeung, C. & Zhang, Z. (2011), 'Self-organization in social tagging systems', *Physical Review E* **83**(6), 066104.

Liu, Z., Lai, Y.-C. & Ye, N. (2002), 'Statistical properties and attack tolerance of growing networks with algebraic preferential attachment', *Physical Review E* **66**(3), 036112.

Lucas Jr, R. E. & Moll, B. (2014), 'Knowledge growth and the allocation of time', *Journal of Political Economy* **122**(1), 1–51.

Lucas, R. E. (1988), 'On the mechanics of economic development', *Journal of Monetary Economics* **22**(1), 3–42.

Lundvall, B.-A. (1992), *National innovation system: towards a theory of innovation and interactive learning*, London: Pinter Publishers.

Luttmer, E. G. (2007), 'Selection, growth, and the size distribution of firms', *The Quarterly Journal of Economics* **122**(3), 1103–1144.

Maddison, A. (2007), *The world economy volume 1: A millennial perspective volume 2: Historical statistics*, Academic Foundation.

Malerba, F. (2002), 'Sectoral systems of innovation and production', *Research Policy* **31**(2), 247–264.

Malerba, F., Nelson, R., Orsenigo, L. & Winter, S. (1999), 'History-friendly models of industry evolution: the computer industry', *Industrial and Corporate Change* **8**(1), 3–40.

Mandelbrot, B. (1953), An informational theory of the statistical structure of language, *in* W. Jackson, ed., 'Communication theory', Betterworth, pp. 486–502.

Mansfield, E. (1961), 'Technical change and the rate of imitation', *Econometrica* **29**(4), 741–766.

Marengo, L. & Dosi, G. (2005), 'Division of labor, organizational coordination and market mechanisms in collective problem-solving', *Journal of Economic Behavior & Organization* **58**(2), 303–326.

Marimon, R. (1997), Learning from learning in economics, *in* D. Kreps & K. Wallis, eds, 'Advances in Economics and Econometrics', Cambridge University Press.

Martínez-Mekler, G., Martínez, R. A., del Río, M. B., Mansilla, R., Miramontes, P. & Cocho, G. (2009), 'Universality of rank-ordering distributions in the arts and sciences', *PLoS One* **4**(3), e4791.

Merton, R. (1968), 'The Matthew effect in science: the reward and communication systems of science are considered', *Science* **159**(3810), 56.

Milgram, S. (1967), 'The small world problem', *Psychology Today* **2**(1), 60–67.

Morris, S. (2000), 'Contagion', *The Review of Economic Studies* **67**(1), 57–78.

Naranan, S. (1970), 'Bradford's law of bibliography of science: an interpretation', *Nature* **227**, 631–632.

Nelson, R. (2006), Perspectives on technological evolution, *in* K. Dopfer, ed., 'The evolutionary foundations of economics', Cambridge University Press, pp. 461–471.

Nelson, R. R. (1959), 'The simple economics of basic scientific research', *The Journal of Political Economy* **67**(3), 297–306.

Nelson, R. R. (1993), *National systems of innovation: a comparative study*, Oxford: Oxford University Press.

Nelson, R. R. & Winter, S. G. (1982), *An evolutionary theory of economic change*, Belknap press.

Neumann, J. v. (1945), 'A model of general economic equilibrium', *The Review of Economic Studies* **13**(1), 1–9.

Newell, A., Simon, H. A. et al. (1972), *Human problem solving*, Prentice-Hall Englewood Cliffs, NJ.

Newman, M. (2009), 'The first-mover advantage in scientific publication', *EPL (Europhysics Letters)* **86**, 68001.

Newman, M. (2010), *Networks: an introduction*, Oxford University Press.

Page, S. E. (2008), *The difference: How the power of diversity creates better groups, firms, schools, and societies*, Princeton University Press.

Pasinetti, L. L. (1983), *Structural change and economic growth: a theoretical essay on the dynamics of the wealth of nations*, CUP Archive.

Paule, P. & Schorn, M. (1995), 'A Mathematica version of Zeilberger's algorithm for proving binomial coefficient identities', *Journal of symbolic computation* **20**(5-6), 673–698.

Pavitt, K. (1984), 'Sectoral patterns of technical change: towards a taxonomy and a theory', *Research Policy* **13**(6), 343–373.

Peltomäki, M. & Alava, M. (2006), 'Correlations in bipartite collaboration networks', *Journal of Statistical Mechanics: Theory and Experiment* **2006**(01), P01010.

Perline, R. (2005), 'Strong, weak and false inverse power laws', *Statistical Science* **20**(1), 68–88.

Peruani, F., Choudhury, M., Mukherjee, A. & Ganguly, N. (2007), 'Emergence of a non-scaling degree distribution in bipartite networks: a numerical and analytical study', *EPL (Europhysics Letters)* **79**, 28001.

Peterson, G., Pressé, S. & Dill, K. (2010), 'Nonuniversal power law scaling in the probability distribution of scientific citations', *Proceedings of the National Academy of Sciences* **107**(37), 16023.

Petkovšek, M., Wilf, H. & Zeilberger, D. (1996), *A=B*, Wellesley, MA, http://www.math.rutgers.edu/ zeilberg/AeqB.pdf.

Polanyi, M. (1962), *Personal knowledge: Towards a post-critical philosophy*, Psychology Press.

Putnam, H. (1975), *Mind, language, and reality*, Cambridge University Press.

Ramasco, J., Dorogovtsev, S. & Pastor-Satorras, R. (2004), 'Self-organization of collaboration networks', *Physical Review E* **70**(3), 036106.

Ratkiewicz, J., Fortunato, S., Flammini, A., Menczer, F. & Vespignani, A. (2010), 'Characterizing and modeling the dynamics of online popularity', *Physical Review Letters* **105**(15), 158701.

Redner, S. (1998), 'How popular is your paper? an empirical study of the citation distribution', *The European Physical Journal B-Condensed Matter and Complex Systems* **4**(2), 131–134.

Riccaboni, M. & Schiavo, S. (2014), 'Stochastic trade networks', *Journal of Complex Networks* **forthcoming**.

Rogers, E. M. (2010), *Diffusion of innovations*, Simon and Schuster.

Romer, P. M. (1990), 'Endogenous technological change', *Journal of Political Economy* **98**(5), S71–S102.

Roth, C. & Cointet, J. (2010), 'Social and semantic coevolution in knowledge networks', *Social Networks* **32**(1), 16–29.

Samuelson, P. A. (1938), 'A note on the pure theory of consumer's behaviour', *Economica* **5**(17), 61–71.

Sanditov, B. (2005), 'Patent citations, the value of innovations and path-dependency', *KITeS Working Papers* (177).

Saviotti, P. (1996), *Technological Evolution, Variety, and the Economy*, Edward Elgar Pub.

Schelling, T. C. (1960), *The strategy of conflict*, Oxford University Press.

Schnettler, S. (2009), 'A structured overview of 50 years of small-world research', *Social Networks* **31**(3), 165–178.

Schultz, T. W. (1961), 'Investment in human capital', *The American Economic Review* **51**(1), 1–17.

Schumpeter, J. (1942), *Capitalism, socialism, and democracy*, New York: Harper.

Schumpeter, J. A. (1934), *The theory of economic development: An inquiry into profits, capital, credit, interest, and the business cycle*, Harvard University Press.

Seglen, P. (1992), 'The skewness of science', *Journal of the American Society for Information Science and Technology* **43**(9), 628–638.

Seidman, S. (1981), 'Structures induced by collections of subsets: A hypergraph approach', *Mathematical Social Sciences* **1**(4), 381–396.

Sen, A. (1999), *Development as freedom*, Oxford University Press.

Shepherd, H. (2010), 'Classification, cognition and context: The case of the World Bank', *Poetics* **38**(2), 134–150.

Shorrocks, A. F. (1975), 'On stochastic models of size distributions', *The Review of Economic Studies* **42**(4), 631–641.

Shy, O. (2001), *The economics of network industries*, Cambridge University Press.

Silverberg, G. (2002), 'The discrete charm of the bourgeoisie: quantum and continuous perspectives on innovation and growth', *Research Policy* **31**(8), 1275–1289.

Silverberg, G. & Verspagen, B. (2005), 'A percolation model of innovation in complex technology spaces', *Journal of Economic Dynamics and Control* **29**(1), 225–244.

Simkin, M. & Roychowdhury, V. (2012), Theory of citing, *in* M. T. Thai & P. M. Pardalos, eds, 'Handbook of Optimization in Complex Networks', Springer Optimization and Its Applications, Springer US, pp. 463–505.

Simkin, M. V. & Roychowdhury, V. P. (2011), 'Re-inventing willis', *Physics Reports* **502**(1), 1–35.

Simon, H. (1955), 'On a class of skew distribution functions', *Biometrika* **42**(3/4), 425–440.

Simon, H. A. (1947), *Administrative Behavior: A Study of Decision-making Processes in Administrative Organization.*, Macmillan.

Simon, H. A. (1956), 'Rational choice and the structure of the environment.', *Psychological Review* **63**(2), 129.

Simon, H. A. (1989), The scientist as problem solver, *in* D. Klahr & K. Kotovsky, eds, 'Complex information processing: The impact of Herbert A. Simon', Lawrence Erlbaum Associates Inc., pp. 375–398.

Simon, H. A. (2002), 'Near decomposability and the speed of evolution', *Industrial and Corporate Change* **11**(3), 587–599.

Simon, H. A. & Ando, A. (1961), 'Aggregation of variables in dynamic systems', *Econometrica* **29**(2), 111–138.

Slater, L. J. (1966), *Generalized hypergeometric functions*, Cambridge University Press.

Soete, L. (2009), 'Malthus' revenge', *UNU-MERIT working paper* **2009-030**.

Solow, R. M. (1956), 'A contribution to the theory of economic growth', *The Qquarterly Journal of Economics* **70**(1), 65–94.

Solow, R. M. (1957), 'Technical change and the aggregate production function', *The Review of Economics and Statistics* **39**(3), 312–320.

Star, S. & Griesemer, J. (1989), 'Institutional ecology, translations and boundary objects: Amateurs and professionals in Berkeley's museum of vertebrate zoology, 1907-39', *Social Studies of Science* **19**(3), 387–420.

Steindl, J. (1965), *Random processes and the growth of firms: A study of the Pareto law*, Griffin London.

Stringer, M., Sales-Pardo, M. & Amaral, L. (2010), 'Statistical validation of a global model for the distribution of the ultimate number of citations accrued by papers published in a scientific journal', *Journal of the American Society for Information Science and Technology* **61**(7), 1377–1385.

Strumsky, D., Lobo, J. & van der Leeuw, S. (2012), 'Using patent technology codes to study technological change', *Economics of Innovation and New Technology* **21**(3), 267–286.

Stumpf, M. & Porter, M. (2012), 'Critical truths about power laws', *Science* **335**(6069), 665–666.

Tsallis, C. & De Albuquerque, M. (2000), 'Are citations of scientific papers a case of nonextensivity?', *The European Physical Journal B-Condensed Matter and Complex Systems* **13**(4), 777–780.

Tsoularis, A. & Wallace, J. (2002), 'Analysis of logistic growth models', *Mathematical Biosciences* **179**(1), 21–55.

USPTO (2012), Overview of the patent classification system, Technical report.

Valverde, S., Solé, R., Bedau, M. & Packard, N. (2007), 'Topology and evolution of technology innovation networks', *Physical Review E* **76**(5), 056118.

Vázquez, A. (2003), 'Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations', *Physical Review E* **67**(5), 056104.

Vega-Redondo, F. (1994), 'Technological change and path dependence: a co-evolutionary model on a directed graph', *Journal of Evolutionary Economics* **4**(1), 59–80.

Vega-Redondo, F. (2007), *Complex social networks*, Cambridge University Press.

Vernon, R. (1966), 'International investment and international trade in the product cycle', *The Quarterly Journal of Economics* **80**(2), 190–207.

Walras, L. (1874), *Éléments d'économie politique pure, ou, Théorie de la richesse sociale*, Corbaz, Lausanne.

Wasserman, S. & Faust, K. (1993), *Social network analysis: Methods and applications*, Cambridge University Press.

Watts, D. & Strogatz, S. (1998), 'Collective dynamics of small-world networks', *Nature* **393**(6684), 440–442.

Weitzman, M. L. (1998), 'Recombinant growth', *The Quarterly Journal of Economics* **113**(2), 331–360.

Youn, H., Bettencourt, L., Strumsky, D. & Lobo, J. (2014), 'Invention as a combinatorial process: evidence from us patents', *arXiv preprint 1406.2938* .

Young, H. P. (2001), *Individual strategy and social structure: An evolutionary theory of institutions*, Princeton University Press.

Young, H. P. (2009), 'Innovation diffusion in heterogeneous populations: Contagion, social influence, and social learning', *The American Economic Review* **99**(5), 1899–1924.

Yule, G. (1925), 'A mathematical theory of evolution, based on the conclusions of Dr. J.C. Willis, F.R.S.', *Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character* **213**, 21–87.

Zeng, A., Yeung, C., Shang, M. & Zhang, Y. (2012), 'The reinforcing influence of recommendations on global diversification', *EPL (Europhysics Letters)* **97**(1), 18005.

Zipf, G. (1949), *Human behavior and the principle of least effort.*, Addison-Wesley Press.

Zlatić, V., Ghoshal, G. & Caldarelli, G. (2009), 'Hypergraph topological quantities for tagged social networks', *Physical Review E* **80**(3), 36118.

Samenvatting

De creatie en verdeling van economische welvaart berust uiteindelijk op de ontwikkeling en verspreiding van kennis. Als gevolg hiervan vereist het begrijpen van de dynamiek, de organisatie en de levensvatbaarheid van de economie een grondige analyse van kennissystemen. Dit proefschrift stelt voor om kennissystemen te bestuderen als zelforganiserende two-mode netwerken. Two mode-netwerken hebben twee soorten knooppunten, en de verbindingen zijn alleen tussen knooppunten van verschillende types. Deze zijn zelforganiserend in de zin dat eenvoudige regels van de evolutie leiden tot een rijke, maar gevormde dynamiek. Het proefschrift bouwt voort op de literatuur over sociale (agent-agent) en epistemische (idee-idee) netwerken om sociaal-epistemische co-evolutie te bestuderen door agent-idee netwerken.

Het blijkt dat: (i) stabiele kracht wet-verdelingen van de populariteit van ideen van nature ontstaat van innovatie en face-to-face diffusie; (ii) deze dynamiek is compatibel met de waargenomen (verschoven) krachtwet van citaties, en (iii) de veralgemeende beta grootte-rank relatie waargenomen voor octrooi-klassen kan worden verklaard door een vertraging in de groei van het aantal klassen. Een algemene les van dit werk is dat kennissystemen vaak dynamica die niet in evenwicht is en niet-lineaire dynamica vertonen, die twijfels kunnen veroorzaken over hun levensvatbaarheid op lange termijn.

Het proefschrift begint met een algemene inleiding (hoofdstuk 1), gevolgd door een overzicht van de literatuur over de kennis en ontwikkeling (hoofdstuk 2) waarin wordt uitgelegd waarom de studie van zelforganiserende kennissystemen een kernonderwerp van de economie zou moeten zijn. Vervolgens wordt een overzicht gegeven van groeiende one-mode en two-mode netwerkmodellen (hoofdstuk 3), samen met enkele originele resultaten. De drie belangrijkste hoofdstukken volgen.

Hoofdstuk 4 presenteert een model van ontwikkeling en verspreiding van ideen in een sociaal netwerk. Agenten leren willekeurige ideen van willekeurige vrienden, daarbij een zichzelf versterkende dynamiek in de diffusie van ideen crerend. Maar deze exponentile diffusie is beperkt aangezien de bevolking begrensd is, wat leidt tot een logistische diffusie-curve. Bij de steady-state, is de verdeling van de populariteit van ideen (het aantal agenten dat een idee kent) een specifieke gegeneraliseerde hy-

pergeometrische verdeling, die de neiging heeft naar de bekende Yule-Simon verdeling wanneer de grootte van de populatie naar oneindig gaat. Een gemiddelde veld zelf-consistentie vergelijking voor de partitie-factor van de bijlage-kernel benadrukt dat een stabiele zelforganisatie plaatsvindt omdat de " overlap 'tussen de agenten' ideen-sets, bepalend en bepaald door welk ideen verspreiden, een stabiele toestand bereikt .

Hoofdstuk 5 gebruikt de resultaten van hoofdstuk 4 om de structuur van citatie-netwerken te verklaren. In hoofdstuk 5 wordt ook aangenomen dat, wanneer nieuwe ideen (papers) verschijnen, ze willekeurige ideen citeren die eerder bekend waren bij de uitvinder. In tegenstelling tot de bestaande modellen, die de exponent van de machtsfunctie citaat-distributies verklaren door de relatieve prevalentie van bibli-ografische (lezen van door andere papers geciteerde papers) en willekeurige zoekop-drachten, toont het model dat de beperkte aandacht en face-to-face diffusie ook de waargenomen patronen kan verklaren.

Hoofdstuk 6 onderzoekt de omvang van octrooi-categorien. De grootte-verdeling van de Amerikaanse octrooi-subklassen is goed uitgerust met een (verschoven) machtsfunctie. Echter, op het niveau van de Amerikaanse octrooiklassen wordt de Zipf-wet duidelijk overtreden en de afmeting-rank relatie is goed uitgerust door een gegeneraliseerde beta-verdeling. Om dit patroon te verklaren, wordt een niet-lineaire aanpassing van het Yule-Simon-Naranan principe voorgesteld. Terwijl volgens laatstgenoemde zowel individuele categorien als het aantal categorien exponen-tieel groeit, wordt hier aangenomen dat het aantal categorien een asymmetrische logistische (Richards)-curve volgt.

About the author

François was born in Clermont-Ferrand (France) in 1984. He obtained an Economic and Social Baccalauréat in 2002, completed a master degree in Economic Analysis and Policy in Clermont-Ferrand in 2007, and obtained a master degree in Industrial Organization and Economics of Knowledge in Strasbourg in 2009. He joined UNU-MERIT in 2009, and is now a research fellow at the Institute for New Economic Thinking at the Oxford Martin School, and at the London Institute for Mathematical Sciences.

2014

85. François Lafond
The evolution of knowledge systems

84. Annalisa Primi
Promoting Innovation in Latin America: What Countries Have Learned (and What They Have Not) in Designing and Implementing Innovation and Intellectual Property Policies

83. Fatoumata Lamarana Diallo
Evaluation of Meal and Deworming Programs for Primary Schools in Rural Senegal

2013

82. Anant Kamath
Information Sharing through Informal Interaction in Low-Tech Clusters

81. Flavia Pereira de Carvalho
What we talk about when we talk about Brazilian Mulitantionals: an investigation on Brazilian FDI, economic structure, innovation and the relationship between them

80. Jun Hou
Complementarity in Innovation and Development: A Cross-country Comparison

79. Rufin Baghana
Impacts of Government Incentives to R&D, Innovation and Productivity:
A Microeconometric Analysis of the Québec Case

78. Lilia I. Stubrin
High-Tech Activities in Emerging Countries: A Network perspective on the Argentinean biotech activity

2012

77. Abdul Waheed
Innovation Determinants
and Innovation as a Determinant:
Evidence from Developing Countries

76. Bilal Mirza
Energy Poverty and Rural Energy Markets in Pakistan

75. Benjamin Engelstätter
Enterprise Software and Video Games: An Empirical Analysis

Fulvia Farinelli
Natural Resources, Innovation and Export Growth: The Wine Industry in Chili and Argentina

Rodolfo Lauterbach
Innovation in Manufacturing: From Product Variety and Labor Productivity Growth to Economic Development in Chile

74. Kirsten Wiebe
Quantitative Assessment of Sustainable Development and Growth in Sub-Saharan Africa.

73. Julio Miguel Rosa
Organizational Strategies, Firms' Performance and Spatial Spillovers. The Canadian Case in Research and Development.

Johannes Wilhelmus Marie Boels
Joseph Schumpeter, honderd jaar economische ontwikkeling. Een historisch-theoretische beschouwing.

2011

72. Daniel Vertesy
Interrupted Innovation: Emerging economies in the structure of the global aerospace industry.

71. Tina Saebi
Successfully managing alliance portfolios: an alliance capability view.

70. Nora Engel
Tuberculosis in India - A case of innovation and control.

69. Evans Mupela
Connectivity and growth in Sub-Saharan Africa: The role of communication satellites

68. Nantawan Kwanjai
Cross cultural intelligence amid intricate cultural webs – A tale of the UnDutchables in the land of 1002 smiles

67. Lina Sonne
Innovation in Finance to Finance Innovation: Supporting pro-poor entrepreneur-based innovation

2010

66. Fernando Santiago
Human Resources Management Practices and Learning for Innovation in Developing Countries: Pharmaceutical Firms in Mexico

65. Zakaria Babutsidze
Essays on Economies with Heterogenous Interacting Consumers

64. Bertha Vallejo
Learning and Innovation Under Changing Market Conditions: The Auto Parts Industry in Mexico

63. Donatus Ayitey
Technical Change, Competitiveness and Poverty Reduction: A Study of the Ghanaian Apparel Industry

62. Sergey Fillipov
Multinational Subsidiary Evolution: Corporate Change in New EU Member States

61. Asel Doranova
Technology Transfer and Learning under the Kyoto regime; Exploring the Technological Impact of CDM projects in developing countries

2009

60. Alexis Habiyaremye
From Primary Commodity Dependence to Diversification and Growth". "Absorptive Capacity and Technological Catch Up in Botswana and Mauritius".

59. Yoseph Getachew
The Role of Public Capital in Economic Development

58. Sandra Leitner
Embodied Technological Change and Patterns of Investment in Austrian Manufacturing

57. Semih Akçomak
The Impact of Social Capital on Economic and Social Outcomes

56. Abraham Garcia
The Role of Demand in Technical Change

55. Saurabh Arora
Coherence in socio-technical systems: a network perspective on the innovation process

2008

54. Rutger Daems
Medicines for the developing world

53. Johannes Hanel
Assessing Induced Technology - Sombart's Understanding of Technical Change in the History of Economics

52. Rifka Weehuizen
Mental Capital: the economic significance of mental health

51. Danielle Cloodt
The relationship between R&D partnership formation, social embeddedness and innovative performance

50. Sabine Fuss
Sustainable Energy Development under Uncertainty

2007

49. Tobias Kronenberg
Reconciling Environmental Conservation with Economic Prosperity: The Feasibility of Double Dividends in the Short and Long Run

48. Viktoria Kravtsova
Assessing the Impact of Foreign Direct Investment in Transition Economies

47. Suhail Sultan
The Competitive Advantage of Small and Medium Sized Enterprises: The Case of Jordan's Natural Stone Industry

2006

46. Bulat Sanditov
Essays on Social Learning and Imitation

45. Mamata Parhi
Dynamics of New Technology Diffusion: A Study of the Indian Automotive Industry

44. Andreas Reinstaller
Social structures and the innovation process: Their role in the demand of firms and consumers

43. Rose Kiggundu
Innovation systems and development : the journey of a Beleaguered Nile Perch Fishery in Uganda

42. Thomas Pogue
The Evolution of Research Collaboration in South African Gold Mining: 1886-1933

41. Geoffrey Gachino
Foreign Direct Investment, Spillovers and Innovation: The Case of Kenyan Manufacturing Industry

40. Önder Nomaler
Technological Change, International Trade and Growth - An Evolutionary, Multi-Agents-Based Modeling Approach

2005

39. Samia Satti Osman Mohamed-Nour
Change and Skill Development in the Arab Gulf Countries

38. Elad Harison
Intellectual Property Rights: Economics and Policy Analysis

37. Daniel Dalohoun
The relationship between R&D partnership formation, social embeddedness and innovative performance: a multi-level approach of social embeddedness

36. Müge Ozman
Networks, Organizations and Knowledge

35. Bas Straathof
Product variety and economic growth - The counteracting effects of scale and idiosyncrasy

34. Wilfred Schoenmakers
Knowledge Flows between Multinational Companies: A Patent Data Analysis

33. Myriam Cloodt
Mergers and Acquisitions (M&As) in High-Tech Industries: Measuring the Post-M&A Innovative Performance of Companies

2004

32. Paola Criscuolo
R&D Internationalisation and Knowledge Transfer. Impact on MNEs and their Home Countries

31. Maarten Verkerk
Trust and Power on the Shop Floor

30. Gottfried Leibbrandt
Adoption, harmonization and succession of network technologies across countries

29. Mark Sanders
Skill Biased Technical change - Its Origins, the Interaction with the Labour Market and Policy Implications

2003

28. Nadine Roijakkers
Inter-firm cooperation in high-tech industries: a study of R&D partnerships in pharmaceutical biotechnology

27. Viki Sonntag
Speed, Scale and Sustainability

26. Masaru Yarime
From End-of-Pipe Technology to Clean Technology

25. Stéphane Malo
The combinatorial Chemistry Revolution - Sustaining a Superior Performance Position through Technological Learning

2002

24. Annelies Hogenbirk
Determinants of Inward Foreign Direct Investment: the Case of the Netherlands

2001

23. John Adeoti
Technology Investment in Pollution Control in Sub-Saharan Africa: The Case of the Nigerian Manufacturing Industry

22. Edward Huizenga
Innovation Management: How Frontrunners Stay Ahead. An Empirical Study on Key Success Factors in the ICT sector

2000

21. Machiel van Dijk
Technological Change and the Dynamics of Industries. Theoretical Issues and Empirical evidence from Dutch Manufacturing

1999

20. Jan Cobbenhagen
Managing Innovation at the Company Level: A Study on Non-Sector-Specific Success Factors

19. Marjolein Caniëls
Regional Growth Differentials: The Impact of Locally Bounded Knowledge Spillovers

1998

18. Aldo Geuna
Resource allocation and knowledge production: Studies in the economics of university research

1996

17. Reinoud Joosten
Dynamics, Equilibria, and Values

16. Hugo Kruiniger
Investment, R&D, and the Financing Decisions of the Firm

1995

15. Hans van Meijl
Endogenous Technological Change: The Case of Information Technology. Theoretical Considerations and Empirical Results

14. René Kemp
Environmental Policy and Technical Change. A Comparison of the Technological Impact of Policy Instruments

13. Rohini Acharya
The Impact of New Technologies on Economic Growth and Trade. A Case Study of Biotechnology

12. Geert Duysters
The Evolution of Complex Industrial Systems. The Dynamics of Major IT Sectors

11. Marjan Groen
Technology, Work and Organisation, A Study of the Nursing Process in Intensive Care Units

1994

10. Huub Meijers
On the Diffusion of Technologies in a Vintage Framework; Theoretical Considerations and Empirical Results

9. Theon van Dijk
The Limits of Patent Protection. Essays on the Economics of Intellectual Property Rights

8. Hans Voordijk
Naar Integrale Logistiek in Bedrijfsketens, Ontwikkelingen in de Bouw

1993

7. Paul Diederen
Technological Progress in Enterprises and Diffusion of Innovations. Theoretical Reflections and Empirical Evidence.

6. Ben Dankbaar
Economic Crisis and Institutional Change. The crisis of Fordism from the perspective of the automobile industry

5. Hanno Roberts
Accountability and Responsibility: The Influence of Organisation Design on Management Accounting

1992

4. Bart Verspagen
Uneven Growth Between Interdependent Economies. An Evolutionary View on Technology Gaps, Trade and Growth

3. Sjoerd Romme
A Self-organization Perspective on Strategy Formation

1989

2. John Spangenberg
Economies of Scale, and Atmosphere in Research Organisations

1988

1. John Hagedoorn
Evolutionary and heterodox innovation analysis : a study of industrial and technological development in process control and information technology