Knowledge Base Diversity, Relatedness and the Spatial Clustering of Innovation

Distance-based Analysis of Patent Activity in Nanotechnology

Universiteit Utrecht

Faculty of Geosciences

Jasper Hoving

MSc Thesis Innovation Sciences

Utrecht University

Supervisor:

dr. Gaston Heimeriks

Second reader:

prof. dr. Koen Frenken

29 January 2015

Preface

This thesis is submitted for the degree of MSc. in Innovation Sciences at Utrecht University. In the master thesis the student has to apply research skills that he or she has learned throughout the academic career. I enjoyed coming up with new solutions to problems that quantitative studies in Innovation Sciences generally face, and trying to theorize about the topic looking from multiple perspectives.

This study looks into the geography of innovation. After following some courses on this subject, it looked like a very interesting topic to dive deeper into. I found out that quantifying spatial clustering of innovation is not as straightforward as I thought beforehand, and that little empirical research with distance-based methods had been done in economics. After a rough start for me personally in the process of writing the proposal, when I started the actual research I enjoyed it very much and the usefulness of this type of analysis really surprised me. In the end I learned a lot from completing a big research project on my own and I gained much knowledge about the field of economic geography.

I would like to express my gratitude to my supervisor Dr. Gaston Heimeriks in particular, as he helped me with valuable feedback and suggestions, and has given me the autonomy in the process to explore new analytical possibilities. Furthermore, I would like to thank my girlfriend Jessy for supporting me and proofreading my thesis.

I hope you enjoy reading this thesis!

Jasper Hoving

Abstract

Empirical evidence has already suggested that both spatial clustering as well as the characteristics of the knowledge base vary over time, but their specific relation has not been researched yet on a global level. The main aim of this thesis is to explore the influence of knowledge base diversity on the spatial clustering of innovation along the life-cycle of the nanotechnology sector. This thesis utilizes a novel approach to globally quantify spatial clustering by calculating a comparable, distance based, clustering index in order to identify evolutionary patterns. USPTO patents with the new 'B82Y' nanotechnology CPC class are used to determine the geography of innovative activity in three areas - the US, Japan and Europe - as well as to calculate diversity and relatedness. A three-year interval is applied in the time-frame of 1990-2011 to capture the early phases of the nanotechnology life-cycle. The results show that the entry- and exit patterns of firms, the amount of patents and the amount of unique geographical locations (cities) are in line with the stylized facts of the life-cycle. From 1990-2005 these variables grew exponentially, after which they stabilize. As expected, in all three areas (the US, Japan and Europe) the degree of spatial clustering rapidly decreases along the cycle, while the degree of knowledge base diversity grows very rapidly in this time-frame. A highly significant negative relation between knowledge base diversity and the degree of spatial clustering was found. This relationship is likely causal, as we observe temporal precedence of knowledge base diversity on the other variables. In the second analysis no relation was found between knowledge base relatedness and the co-localization of nanotechnology fields. However, the difference in knowledge base diversity between fields does influence their co-location patterns, meaning that fields with lower diversity co-locate with fields that have a higher knowledge base diversity. Furthermore, nanotechnology fields with similar degrees of related variety tend to co-locate as well. The third analysis shows that both nanotechnology diversification and related variety inside nanotechnology clusters positively influence the relative innovative performance of these clusters significantly. The general conclusion from this thesis is that not only regional knowledge base parameters influence spatial clustering, but that throughout the early life-cycle phases of the nanotechnology sector the changing knowledge base characteristics, and in particular diversity, influence both the propensity to spatially cluster as well as the dispersion of geographical locations on a global level.

Contents

CHAPTER 1

Introduction

The spatial economic landscape is uneven, showing spikes where economic activity is geographically concentrated (Florida, 2005). And even though costs of information, communication and transportation are diminishing, spatial agglomeration forces are, paradoxically, increasing in importance (Groot et al., 2007). A key issue in economic geography is whether and in what way firms benefit from proximity to other firms and institutions. In other words, why do firms tend to cluster in space?

When looking at the regional economy from an evolutionary perspective, technological change, knowledge and innovation play a central role as they have the capability to generate new economic opportunities (Crescenzi, 2014). Looking beyond the neo-classical economic models, knowledge and innovation are inherently embedded in economic growth. The knowledge spillover is an important concept in innovation sciences as the R&D activities that firms undertake to advance technology do in general not proceed in isolation, but they are supported by various external sources of knowledge (Breschi, 2000). The effects of knowledge spillovers decline with distance and therefore innovation is considered to be localized (Malecki, 2014). In other words, for innovative activity itself "geography matters" (Audretsch & Feldman, 2004) and knowledge production is often path- and place dependent (Heimeriks & Boschma, 2013). This results in a non-homogeneous geographical pattern of innovation; it is spatially clustered (Malmberg et al., 1996).

We know from empirical research that the propensity to cluster for firms differs along the life-cycle of an industry or sector (Audretsch & Feldman, 1996b; Dumais et al., 2002). Given the research on knowledge spillovers, the nature of a firm's knowledge base strongly affects their innovative opportunities and the their ability to get access to knowledge externalities (Breschi, 2000). As Cowan et al. (2004) put it: the value of clustering is "a function of the characteristics of the knowledge base". It is believed that these characteristics of the knowledge base of a sector or industry influence the spatial patters of innovation and that both the degree of spatial clustering and these characteristics are subjected to change over time. For instance, it is argued that the amount of tacit knowledge and the complexity of technical knowledge involved in the industry determine the need for geographical proximity (Howells, 2002; Sorenson et al., 2006).

In this study one tangible aspect of the knowledge base, its diversity, is being researched. The diversity of the knowledge base is likely intertwined with the development of both tacit knowledge and knowledge complexity. As the understanding of core technologies of a sector increases, and thus its complexity decreases, knowledge codification rises and new application areas open up. Firms can only assimilate and apply knowledge that is related to their own knowledge base (Cohen & Levinthal, 1990), explained by the concept of absorptive capacity and cognitive proximity (Nooteboom, 2000), which in turn can be facilitated more easily by geographical proximity. New application areas and new technologies involved in the knowledge base, resulting in a higher diversity of the field, may thus influence the need for geographical proximity.

Studies have already shown that properties of industries in regions, such as diversity (Duranton & Puga, 2001), related variety (Frenken et al., 2007) and relatedness (Rigby, 2013), are relevant for the performance and production of these regions or cities. Although these results provide valuable information about the locational choices, entry and birth of firms, they do not explain why spatial patterns of innovation differ in the evolution of an industry or sector. In other words, it can be useful to look at it another way, using the characteristics of the knowledge base as an explanans for spatial concentration. To my knowledge, the relation between the development of knowledge base diversity and the global patterns in spatial clustering of innovation over time has not yet been studied.

The sector that is researched in this thesis is nanotechnology. This set of technologies originated in the 1960's, while really gaining momentum in the early 1990's. This sector is chosen for two reasons. First, nanotechnology is a complex technological field which depends greatly on the R&D within firms and institutions. This makes knowledge and innovation likely an important factor in the geographical location of firms, which is were I am interested in. Second, nanotechnology is an attractive sector to study as it is regarded as a generic technology, which can lead to value creation in many domains (Maine & Garnsey, 2006). It is a multidisciplinary field with research and products in for example materials, biochemistry, optics and computer science. These different sub-fields are useful for analyzing the development of the knowledge base of nanotechnology.

The main aim of this thesis is to find the relation between knowledge base diversity of nanotechnology and the spatial clustering of innovation in this sector. I thus ask the question:

What is the influence of knowledge base diversity of nanotechnology on the spatial clustering of innovative activity in this sector?

Essential when using an evolutionary framework is to analyze the topic with temporal data. Nanotechnology is therefore studied over time in the period from 1990 to 2011. It is interesting to see how the patterns of knowledge base diversity and spatial clustering relate to the existing theories on the life-cycle of a sector. Although a stylized concept, the life-cycle provides a good explanation for the different behavior of innovative activity over time. Life cycle dynamics are in literature mainly described by the entry- and exit patterns of firms along the cycle and the diffusion of innovation, which follow an s-shaped curve (Andersen, 1999). As nanotechnology is in the early phases of the cycle (introduction and growth phase), it is expected that the number of entrants will significantly increase in these early phases. However, we do not know what this does to the number of geographical locations (cities) of where innovative activity takes places and how spatial clustering behaves along the cycle. The first sub-question is therefore:

SQ1: How does the number of geographical locations and the degree of spatial clustering of innovation in nanotechnology evolve along the life-cycle?

Then, relating to the first question and in order to find the connection between spatial clustering and knowledge base diversity, I ask:

SQ2: Does knowledge base diversity influence the spatial clustering of innovation in nanotechnology?

Besides analyzing spatial clustering of nanotechnology in general, another perspective on the geographical dynamics can be gained by exploring the co-localization of different nanotechnology fields. This study researches 7 nanotechnology sub-fields and tries to find out whether these fields co-locate in geographical space. The aim is to find out how the relatedness of the knowledge bases of the nanotechnology fields is connected to the tendency to locate near each other. Therefore I ask:

SQ3: Does the relatedness of knowledge bases influence the co-localization of nanotechnology fields?

Lastly, also nanotechnology clusters and their performance are studied. Theoretically, the diversification and the related variety in nanotechnology fields inside the clusters should impact the performance of these clusters. Therefore, the third sub-question is posed as follows:

SQ4: Does the diversification and related variety of nanotechnology in clusters influence their relative performance?

In order to answer these questions, USPTO patent data is used as a proxy for innovative activity. About 39.000 nanotechnology patents are gathered by using the new 'B82Y' CPC class and spatial clustering is calculated through time in the US, Japan and Europe. This research applies a relatively new approach to quantify spatial clustering, namely the k-density function developed by Duranton & Overman (2005). This kind of distance-based measurement is more common in ecology and epidemiology studies, but is still scarcely applied in economics. This research applies the distance-based measurement of spatial clustering in a novel way; a clustering index is calculated over time and related to independent variables in a regression. This study thus provides new empirical and methodological results that show the usefulness of distance-based measurement in economic and innovation studies. Another scientific contribution is that the knowledge base diversity in sectors and technology fields has not yet been studied in relation to the global patterns of spatial clustering over time. The results from this analysis provide new insights in the industry life-cycle as well. Furthermore, the co-localization of technology fields has not been previously linked to the relatedness in knowledge bases. The goal is to contribute to our understanding of how knowledge base characteristics influence the co-location of innovation in different technology fields. Lastly, the effects of diversification and related variety in clusters have not yet been researched for nanotechnology and for innovative activity in general.

There are also some societal contributions made by this study as well. The results can have implications for policy makers as regional and national governments often struggle with how to create a hospitable environment for high tech firms. The results from this thesis can complement our existing knowledge of regions as the clustering patterns related to the life-cycle phase of a high-tech sector provides information about the timing and the possible effect of policies. Furthermore, there can be implications for investments - made by high-tech firms themselves and by governments - which are influenced by geographical context. Knowledge about the influence of diversity in knowledge bases on clustering patterns can be helpful in dealing with these kinds of questions as they can improve the choices for the best geographical location for investments (e.g. location of R & D centers, science parks etc). Besides the general insights about spatial clustering, we also gain knowledge about the nanotechnology sector. The information about the co-location of technology fields and performance of nanotechnology clusters can inform both policy makers (e.g. the European Union) and firms about the state of nanotechnology in their region. An example for The Netherlands could be the performance of the Eindhoven nanotechnology cluster in comparison to the rest of Europe.

This thesis begins first in chapter 2 with explaining the theory and previous studies on this topic. In this chapter also the hypotheses are posed that will be tested. Following this, in chapter 3 the data is described as well as some general descriptive statistics about the database. In part I of the analysis, chapter 4, first the methods for calculating spatial clustering and knowledge base diversity are described. Then, the results from the analysis are elaborated upon. Chapter 5 contains the methods and results for the co-localization of the nanotechnology fields. Part III of the thesis describes the methods and results for the analysis of cluster performance in chapter 6. I discuss the results and the reliability and validity of the research in chapter 7. Finally, the conclusions are drawn in chapter 8.

CHAPTER 2

Theory

2.1 Economic Geography and the Geography of Innovation

Before the rise of the New Economic Geography (NEG), introduced by Krugman's seminal paper from 1991, spatial economics only comprised a relatively small part of mainstream economics[1]. The classical works from for example Von Thünen (1826) and Weber (1909) were based on the idea to identify the regularities of the neoclassical space-economy (Scott, 2000). The model of Krugman was important in that it provided scholars with formal modeling tools for explaining the differences between regions and it caused a renewed attention from economists in regional sciences. Although arguably simplistic from a geographer's point of view, one of the main advantages of Krugman's core-periphery model is that it allows for convergence or divergence between regions as opposed to the neoclassical model (Gaigné & Thisse, 2014).

Paradoxically however, and despite the fact that Krugman and followers have named their approach the 'New Economic Geography', the writings of Krugman received little attention of economic geographers (Martin & Sunley, 1996). While the usefulness of the formal modeling is apparent in describing the economics of regional sciences, it still remains a neoclassical approach with the use of equilibriums. Therefore in recent years the 'Evolutionary Economic Geography' (EEG) has tried to combine the previous works of evolutionary economists, such as Nelson & Winter (1982) and Dosi & Nelson (1994), with elements from economic geography. The goal is to develop a general theory of economic geography which can be sufficiently corroborated by empirical evidence (Boschma & Frenken, 2011b). In EEG the context and environment of economic change is key to be able to understand the processes involved; it is important to analyze how the landscape evolves over time (Boschma & Martin, 2010). One of the most important implications of EEG is that the region is "conditioned by its own technological development trajectory": its own technological history (Mccann, 2014). The framework of EEG is rather complex as multiple levels must be analyzed to

[1] See Scott (2000) for a full historical review of economic geography in the 20th century

understand how the system works. At the micro-level organizational routines are the unit of analysis, based on the work of Nelson & Winter (1982), explaining the decision making (entrepreneurship) under bounded rationality (Boschma & Frenken, 2006). On the meso-level both the spatial evolution of sectors and networks are relevant. At the macro-level structural change is observed as an aggregate of sectors and networks (Boschma & Frenken, 2006). EEG theory is relatively early in its stage of development (Bathelt & Peng-Fei, 2014) and more empirical evidence is needed to be able to get a better view of the complex processes involved. In this thesis I will work from an evolutionary perspective on the meso/macro-level, analyzing the spatial patterns of innovation in the nanotechnology sector.

Innovation and knowledge take a central position in economic geography when looking from an evolutionary perspective (Boschma & Martin, 2010). Technological advance and the accumulation of knowledge are considered to be main explandi for national and regional differences in economic growth (Nelson & Winter, 1982; Romer, 1986). In this sense, economic growth can be modeled as induced changes in knowledge, technology, organization and location (Karlsson & Grå sjö, 2014). As discussed before, economic activity is geographically discontinuous. Previously, this lumpiness was attributed to localized natural resources, capital investments and entrenched relationships between firms and industries. However it is increasingly viewed as also being caused by the 'spiky' development of knowledge in geographical space (Rigby, 2013).

Already in 1890 Marshall argued that knowledge can accumulate in a region and provide economic externalities to firms in that region. Studies from the early 1990's (e.g. Jaffe, 1989; Acs et al., 1992) have shown that produced knowledge can spill over from firms and research institutions for third-parties to exploit. As argued by many, tacit knowledge is a key aspect in the geography of innovation (Asheim & Gertler, 2005). Tacit knowledge is non-codifiable, context dependent knowledge, which involves social interaction in order to transfer it. This makes it more costly or difficult to transfer over greater distance thus resulting in the importance of proximity for knowledge transfers. Thus, tacit knowledge is 'sticky' in respect to the location it is produced. Empirical research has shown that R&D intensive industries are more spatially concentrated, which suggests that localized spillovers play a role and not merely the geographic concentration of production (Breschi, 2011). Even for codified knowledge geographical proximity might be of influence as the interpretation and assimilation may still contain a tacit component (Howells, 2002; Boschma, 2005).

A second aspect of this place-dependence is that innovative activity itself is becoming increasingly complex and more dependent on social interactions between firms, research organizations and public agencies (Asheim & Gertler, 2005). Malmberg & Maskell (2010) coined for this the term 'localized learning', which comprises a horizontal dimension (firms in related industries), a vertical dimension (firms up and down in the supply chain) and a social dimension (neighborhood effect). Complexity drives the need for knowledge exchanges in these three dimensions, involving a certain degree of geographical proximity.

Both tacit knowledge and localized learning thus show that 'geography matters' for innovative activity. However, this does not mean that proximity is prerequisite for firms to interactively learn from each other (Oinas & Malecki, 2002). In fact, social networks do not have to be localized geographically as "there is nothing inherently spatial about networks" (Boschma, 2005). Carrincazeaux & Coris (2011) argue that agglomeration alone does not generate cooperation between actors but its association with other forms of proximity.

2.2 Agglomeration externalities and spin-off dynamics

A classical issue in economic geography is what the effects of geography are on the performance of firms (Neffke et al., 2011). It is recognized that differences exist between the benefits that firms receive depending on the region or city. These effects are often labeled spatial or agglomeration externalities. In turn, the externalities where firms benefit from influence the spatial clustering of economic and innovative activity itself. These effects are often referred to as agglomeration forces. The interplay between externalities and agglomeration forces are combined in the concept of agglomeration economies. Although it is generally accepted that various sources of agglomeration externalities can be distinguished, the literature is not unambiguous in these distinctions. In this thesis I adopt the externalities used in the review study from Beaudry & Schiffauerova (2009):

1. *MAR (Marshall-Arrow-Romer) externalities*. The specialization of an industry in a region promotes knowledge exchange between firms and institutions resulting in an increase of intra-industry knowledge spillovers.

2. *Jacobs externalities*. Diversity of industries in a region promotes knowledge externalities resulting in an increase of inter-industry knowledge spillovers.

3. *Porter externalities*. Local competition in a region provides firms with an incentive to innovate, resulting in productivity growth. Spillovers occur vertically within the industry.

As Beaudry & Schiffauerova (2009) points out, Porter externalities resemble elements from both MAR externalities (intra-industry spillovers) and Jacobs externalities (competitive environment). The main question is thus whether MAR or Jacobs externalities benefit firms, as they seemingly contradict each other. Although this topic has been researched extensively there is no conclusive answer to this question. The meta study by Groot et al. (2007) concludes that there is evidence for Jacobs externalities but no effects are found for MAR. However, in their review paper Beaudry & Schiffauerova (2009) find that both types of externalities influence agglomeration. Boschma & Frenken (2011a) argue that from an evolutionary perspective it might be good to move beyond ‘MAR vs Jacobs’. In recent approaches a distinction in Jacobs externalities is made between related and unrelated variety (Frenken et al., 2007). It is argued that knowledge spillovers cannot happen between all types of actors; their cognitive distance must be not too small and not too large. Nooteboom (2000) theorizes that while a knowledge exchange with great cognitive distance has a high degree of possible novelty, it also has the problem of incomprehensibility. Therefore a trade-off should be made between cognitive proximity and cognitive distance. With this theory in mind, Boschma & Frenken (2011a) argue that not diversity per se has to be a factor in agglomeration economics, but that there should be a degree of relatedness between technologies. A distinction is made between ’related variety’ and ’unrelated variety’. As high-tech innovation comprises a recombination of knowledge, related variety is expected to be beneficial for firms. If related industries are missing in the region (unrelated variety) firms cannot benefit from knowledge spillovers. Although relatively little empirically tested, the distinction in unrelated/related variety seems to produce somewhat more consistent results than taking Jacobs externalities as a whole (see for example Boschma & Iammarino (2009) who test also for Jacobs). Continuing on the theory of related variety, the following two hypotheses are devised:

Hypothesis 1: The relative performance of nanotechnology clusters is positively influenced by the degree of diversification of nanotechnology fields inside the clusters.

Hypothesis 2: The relative performance of nanotechnology clusters is positively influenced by the degree of related variety within the nanotechnology knowledge base inside the clusters.

Hypotheses 1 and 2 are clearly linked with each other. Diversification in hypothesis 1 is defined as the variety of technology fields *within* the geographical cluster itself, so for example a cluster that has equal innovative activity in both nano-optics and nano-magnetism is considered more diversified than one where innovation is dominated only by nano-optics. The possitive effects of industry diversification on cluster performance have already been shown by Maggioni (1999), but it is interesting to see whether the same results are obtained for innovative activity.

When a cluster is more diversified, the related variety within the nanotechnology knowledge base [2] of the cluster is also expected to increase [3]. Therefore it is expected that both

[2]It must be noted that related variety is not used in exactly the same way as Frenken et al. (2007) defined the concept. Instead of calculating industry relatedness, the relatedness of the knowledge base is used. This is necessary as this thesis is only exploring one industry, namely nanotechnology. Nevertheless, I expect it represents the same concept, as the relatedness in innovative industries mainly depends on their knowledge base. More about the concept of the knowledge base and technological relatedness is discussed in section 2.4

[3]It could also be argued that an increase in related variety causes the opening of new technological possibilities, resulting in a diversification of the cluster in new (related) technological fields. This possible causal relationship and its direction is not being researched in this thesis

diversification and related variety of the knowledge base are positively related to the relative performance of the cluster.

Besides agglomeration economies, there is another force that drives spatial clustering: spin-off dynamics. As shown by Klepper (2007), better performing firms produce more spinoffs, which often locate in the same area as their parent. This Darwinian evolutionary process can thus cause a self-reinforcing clustering pattern without the influence agglomeration economies involved. Although Klepper (2007) and Buenstorf & Klepper (2009) show that the evolution of the U.S. car industry and the Detroit agglomeration was under influence of spin-off dynamics and that the effect of agglomeration economies was minimal, Boschma & Wenting (2007) find that both of these forces were important regarding firm location. It is likely that the influence of agglomeration economies and spin-off dynamics on spatial clustering differs between types of industries or technologies as well as the stage of the life cycle the industry is in.

2.3 Spatial Clustering and Life-cycle patterns

First, it is important to define what is meant by the concept of spatial clustering. Previously, I discussed geographical proximity, which simply refers to "the spatial or physical distance between economic actors, both in its absolute and relative meaning" (Boschma, 2005). When looking at a group of points in geographical space, variances in distance between points can result in a non-homogeneous pattern. Statistically, if a set of spatial points lacks homogeneity the set is considered spatially clustered. A cluster can thus be defined as a non-random spatial concentration of economic activity (Ellison et al., 1997), or in this case, innovative activity. Although spatial clustering is a multi-dimensional concept (Steinle & Schiele, 2002), in this research it is used as a statistical concept to quantify the geographical concentration of innovation in a geographical area. Spatial clustering can be higher or lower in degree, or in other words; more localized or more dispersed. In this thesis the notion of spatial clustering is used instead of concepts such as *localization* and *industry agglomeration*, in accordance to Malmberg et al. (1996) and Malmberg & Maskell (2002).

An interesting issue is whether the clustering of innovation also follows a specific pattern during the life-cycle. Although the life-cycle model is highly stylized and not all industries or technologies necessarily follow the same pattern (Boschma & Martin, 2010), there are some theoretical arguments to make why geographical clustering is also linked to the industry life-cycle. First, because the intensity and type of innovative activity varies in time, and as knowledge spillovers are influenced by the extend to which tacit knowledge is present, the geography of innovation is likely affected by the life-cycle stage. Second, as the entry- and exit of firms changes along with the developments in the industry, this has consequences for the clustering of these firms and therefore the clustering of innovation. The life-cycle has been used often by economic geographers to help explain spatial clustering (Boschma &

Martin, 2010). I will highlight four stylized aspects of this cycle that are relevant for spatial clustering.

First, among Schumpeterian economists it is commonly argued that development and diffusion of innovation follows an S-shaped function (Andersen, 1999). One of the most studied aspects of the industry life-cycle are the entry and exit patterns of firms. A more or less fixed pattern is found among industries, in which the industry formation is characterized by significant entry of firms, then leveling off and begins to decline, even before maturing (Audretsch & Feldman, 1996b). Another characteristic of this cycle is that innovative activity tends to be highest in the early phases, while in the maturing phases product innovation decreases and process innovations become priority (Klepper, 1997).

Second, the formation of geographical clusters at the start of the life-cycle is rather obscure and the clustering process is largely random with respect to location (Press, 2006). Of course, some geographical places with pre-existing resources relevant to the industry or sector will be more likely to develop into a cluster, but in general it is almost impossible to determine ex-ante where clusters will emerge (Malmberg & Maskell, 2010). The paradox here is that a cluster is itself is defined as a non-random concentration of activity and the roots of a cluster can often be traced by an ex-post analysis (Malmberg & Maskell, 2010).

Third, it is generally accepted that the propensity of tacit knowledge in an industry or sector varies along the life-cycle. As there is more uncertainty and there are no standards available in the industry in the early phases of the cycle, tacit knowledge plays an important role here (Audretsch & Feldman, 1996b). Information is less codified than in mature industries, and face-to-face communication is a more common way of information transmission (Cowan et al., 2004). When the industry matures the products and technologies are better understood and information becomes more codified, illustrated by the fact that in these phases of the cycle the amount of imitators in the market increases. The role of tacit knowledge, and thus the life-cycle phase, is linked to spatial clustering as tacit knowledge transfer costs increase with distance. Furthermore, the complexity of knowledge decreases along the life-cycle, and the advantage of actors having access to a template of complex knowledge diminishes as an industry progresses, reducing its geographical concentration (Sorenson et al., 2006). The evolution of tacit knowledge and complexity then results in a lower propensity to cluster in mature phases of the cycle in comparison to the introduction and growth phase of the cycle, where clustering is more beneficial for innovative activity.

Fourth, evolutionary theory predicts that regional specialization depends on the life-cycle phase of the industry (Boschma & Wenting, 2007). Most scholars agree about the theory that Jacobs externalities are a stronger force in the early phases of life cycle. This because in the early phases of the cycle the innovative activity (and uncertainty) of firms in product development is high, requiring a variety of knowledge. The need for a recombination of existing knowledge to form new products causes a tendency for firms to locate near each other to gain benefits from a diverse environment. When the industry matures, innovation becomes less radical and the focus is shifted to incremental innovations. The need for a more low-cost supply chain increases and the tendency to specialize becomes greater. Thus, the proximity to actors in the same sector becomes more important, enhancing the benefits from MAR externalities. These patterns of Jacobs in the early phases and MAR in later stages have indeed been found by Duranton & Puga (2001) and Neffke et al. (2011). However, it is still unclear how this plays a role in a technology or sector in general, as these studies mostly only look at regions. Furthermore, little evidence exists of how externalities and spin-off dynamic behave as the life-cycle progresses.

The stylized facts described above have important implications for the spatial clustering along the life-cycle. Innovation in nanotechnology in the observed time-frame likely represents the early stages of its life-cycle. Therefore, it is expected that there is a significant entry of new firms and research institutions during this period, like Klepper (1996) predicts and many empirical research has shown (e.g. Utterback & Suárez, 1993; Horvath et al., 2001; Agarwal & Audretsch, 2001). The great increase of firm entry and thus increasing innovative activity is expected to also increase the amount of geographical locations where this activity takes place. Therefore, the third hypothesis is stated as follows:

Hypothesis 3: The number of geographical locations increases in the early phases of the nanotechnology life-cycle while stabilizing later on in time.

This increase of geographical locations has its effects on spatial clustering. Dumais et al. (2002) find that the opening up of new firm locations throughout geographical space decreases the degree of spatial clustering in most industries they study. Although Audretsch & Feldman (1996b) and Dumais et al. (2002) study plant location and not innovation like in this thesis, I expect to find similar results. Spatial clustering is expected to be high at the start of the cycle, while decreasing over time. Therefore, I hypothesize:

Hypothesis 4: The degree of spatial clustering of innovation decreases in the early phases of the nanotechnology life-cycle while stabilizing later on in time.

2.4 Nanotechnology and knowledge base diversity

2.4.1 The knowledge base of Nanotechnology

Nanotechnology is recognized for its important future applications and the possible great impact on society it can have. It comprises the imaging, measuring, modeling and manipulating of matter at a scale of 1-100 nm (Lu, 2009). Nanotechnology is an umbrella term for a range of technologies and stem from basic research in the field of nanoscience, which represents a convergence of quantum physics, molecular biology, computer science, chemistry and engineering (Mehta, 2002). Because of this convergence and the different application domains (e.g. materials, biochemistry, optics) nanotechnology is considered to be a multidisciplinary field to a great extend. The concept of nanotechnology originates from a speech by Richard Feynman in 1959 and the term nanotechnology is first used by Taniguchi & Others (1974). In 1981 Gerd Karl Binnig develops the scanning tunneling microscope (STM) which provided the first images of individual atoms on material surfaces (Rothaermel & Thursby, 2007). Nanotechnology really gained momentum in the early 1990's with the invention of the atomic force microscope (AFM), which allowed researchers to examine non-conducting material surfaces.

From the start, nanotechnology is seen as one of the scientific fields having great technological and economic potential with possible breakthroughs in medicine, manufacturing, high-performance materials, information technology, and energy and environmental technologies (Rothaermel & Thursby, 2007). Nanotechnology is considered as a generic technology (Asheim et al., 2011) and thus has "the potential for value creation across a broad range of industries and applications" (Maine & Garnsey, 2006). In a generic technology knowledge spillovers are an important phenomenon as knowledge is exchanged between the related industries (Frenken et al., 2007; Asheim et al., 2011).

Nanotechnology is of today often still regarded as an emerging set of technologies, although the amount of nanotech products appearing in the market is growing relatively fast (Shapira et al., 2011). In fact, Shapira et al. (2011) provide evidence that nanotechnology is shifting from the discovery phase to commercialization since 2003. The nanotech consumer products now on the market can be considered as incremental improvements to existing products, such as nanosilver particles. Since the 1990's there has been a steady increase in the amount of firms active on the market, and there are many start-ups that spin-off from universities. Rothaermel & Thursby (2007) provide an interesting comparison of the evolution of biotechnology and nanotechnology and conclude that large incumbent firms were important from the start in nanotechnology, contrasting the early phases of biotechnology in which start-ups were dominant. Rothaermel & Thursby argue that this may be the result of the quick technological succession of the STM by the AFM.

2.4.2 Knowledge base diversity

Asheim et al. (2011) distinct three different types of knowledge bases: analytical (science based), synthetic (engineering based) and symbolic (arts based). As nanotechnology is a generic, research based industry, its knowledge base can best be characterized as analytical. An analytical knowledge base implies that knowledge is codified and as a result that knowledge creation is less sensitive to distance-decay (Asheim et al., 2011). However, this certainly does not mean that firms do not cluster in space. Quite the contrary is true, as can be seen for example in biotechnology and the IT-sector. Particularly, the role of inter-industry knowledge spillovers results in an important role for geography in high-tech industries (Audretsch & Feldman, 1996b; Hartog et al., 2012).

The main goal of this thesis is to research the relation between the diversity of the nanotechnology knowledge base and the spatial clustering of innovation. The connection between geography and the knowledge base has been researched before, but is mostly concentrated on firm-level knowledge base, regional branching, cities (clusters) and relatedness. The influence of the diversity the knowledge base of a technology or sector on clustering in general has been barely touched upon by scholars. And there is a good theoretical basis to assume that there is a relation between the two, or as Cowan et al. (2004) put it: the value of clustering is "a function of the characteristics of the knowledge base".

First, in recent years there has been an increasing interest in the regional knowledge base and the influence on firm entry as well as cluster- or regional performance. Looking into regional innovation systems (RIS), Asheim & Coenen (2005) discovered that the type of knowledge base of a sector can influence the needs for the type of infrastructure and knowledge mechanisms in regions. Simmie & Martin (2010) found with the help of case studies that the endogenous knowledge base of a region may well be one of the key factors for the economic resilience of a region. Furthermore, as discussed before, the research into externalities and related variety have pointed out that the knowledge base in a region is of influence on the performance and entry of firms in the region. The notion that the regional knowledge base is important gives way to assume that the relation also exists the other way around; the knowledge base of a sector or technology field influences the geographical dynamics of firms and innovation in that sector.

Second, the nature of a firm's knowledge base strongly affects their innovative opportunities and the their ability to get access to knowledge externalities (Breschi, 2000). The concept of absorptive capacity (Cohen & Levinthal, 1990) and cognitive proximity (Nooteboom, 2000) are both dependent on the knowledge bases of firms. In turn, these two concepts likely influences the geographical dynamics of innovation in general and can drive agglomeration forces. For instance, McCann & Folta (2011) have found that firms with a deeper knowledge base, thus having a higher absorptive capacity, benefit more from being located in knowledge-rich clusters. The knowledge base of firms thus seems to determine the spatial boundaries in which they can search for knowledge inputs.

Thirdly, Breschi (1999,2000) shows that spatial patterns of innovation among technological classes and sectors differ systematically, meaning that this can be attributed to industry- or technology-specific factors. In other words, the knowledge base of a sector or technology class seems to influence the spatial patterns of innovation. In the typology composed by Carrincazeaux & Coris (2011) (table 2.1) the need for geographical proximity depends on the technological complexity (defined as the renewal frequency of the knowledge base) and the combinatorial complexity (reflecting the difficulty of consistency of diverse pieces of knowledge) (Carrincazeaux & Coris, 2011). Although it highlights the influence of the knowledge base on the need for proximity, Carrincazeaux & Coris (2011) stress that "geographical proximity is not in itself a factor of coordination: insufficient in itself, it must be enabled by the existence of organizational and institutional proximity". Although the nanotechnology knowledge base probably can be characterized by both a high combinatorial and technological complexity, it might well be the case that the need for geographical proximity varies over the life-cycle of nanotechnology as factors such as uncertainty and tacit knowledge may influence the need for organizational and institutional proximity as well. However, not much is known about these dynamics as Boschma (2005) also points out.

Table 2.1: Forms of proximity and dominant spatial configurations according to knowledge base complexity. Adopted from Carrincazeaux & Coris (2011)

		Technological complexity	
		Strong	Weak
Combinatorial complexity	Strong	Geographical proximity *Agglomeration, clusters*(1)	Organizational proximity *Nomad organization/transitory*(4)
	Weak	Institutional proximity *Reticular organization/ temporary* (3)	Weak need for proximity *Distant interactions*(2)

Breschi argues that the more tacit knowledge and complexity is involved, the more geographically concentrated innovative firms will be. As was argued in the previous section, the amount of complexity and tacit knowledge decreases with the advancement in the life-cycle, especially when entering the growth phase. I expect that when the complexity and tacit knowledge decrease, the diversity of the knowledge base is likely to increase, at least in the first phases of the cycle. When the core technologies that supported the nanotechnology sector from the start mature and uncertainty about the potential of these technologies decrease, the understanding and codification of knowledge will increase. This pattern, predicted by the life-cycle theory, results then in the emergence of new fields and application areas related to these core technologies, which will receive increasing attention of firms and institutions. The scope of innovation in nanotechnology thus diverges, leading to an increase in the diversity of the knowledge base of the sector. The increasing diversity attracts new firms (startups or incumbent firms from related fields) and research facilities to venture in nanotechnology and opens up new geographical locations. This early growth in the life-cycle of new locations caused by an increasing technological diversity is then expected to lead to a decrease of spatial clustering of innovation in nanotechnology. Therefore, the hypothesis I pose is:

Hypothesis 5: The degree of knowledge base diversity negatively influences the degree of spatial clustering of innovation in nanotechnology in the first phases of the life-cycle.

As part of the spatial clustering of nanotechnology as a whole sector, the technology fields within nanotechnology (e.g. nanomechnatism, nanobiotechnology) can also be studied.

Interesting here is to analyze how the knowledge base of these technology fields influence their co-location in geographical space. In literature, the degree to which industries, sectors or firms co-locate is referred to as ’co-localization’ (also: ’co-agglomeration’). Studies have pointed out that there indeed is a tendency of co-localization of firms between certain sectors (Ellison et al., 1997; Duranton & Overman, 2005; Braunerhjelm & Borgman, 2006; Arbia et al., 2007). Some of the studies (Ellison et al. (1997) and Braunerhjelm & Borgman (2006)) link the constructed co-localization indices to variables such as labor pools and productivity, resulting in both evidence for the influence of MAR and Jacobs externalities. These results imply that the co-localization of technology fields may be related to the knowledge bases of these fields. Furthermore, research has revealed that technological relatedness affects the knowledge space and the development of regions (Rigby, 2013; Boschma et al., 2014). The development of new technology fields depends on a re-combinatorial process of existing technologies in the knowledge space and new fields generally are more likely to develop in a specific region when related fields are already present in that geographical area. In the same analogy, we can assume that the more related two nanotechnology fields are in terms of their knowledge base, the more likely they will be co-agglomerated geographically. Therefore, the last hypothesis of this thesis is formulated as follows:

Hypothesis 6: The degree of the co-localization of two nanotechnology fields is positively related to the degree of relatedness of their knowledge bases

CHAPTER 3

Data

3.1 Patent Data

Patent data is used in this thesis in order to assess the spatial clustering of innovation in nanotechnology. Patents are legal instruments in economic life and used to protect inventions (OECD, 2009). Many studies on measuring innovative activity have been in the form of patent analyses, as they provide a vast amount of data about these inventions. Griliches (1998) puts it: "In spite of all the difficulties, patent statistics remain a unique resource for the analysis of the process of technological change".

In general, patents have some advantages over other science & technology data. First, patents are regarded as an outcome of the innovative process, in contrast to R&D investments which measure innovative input of firms. Studies point out that there is a relationship between patent output and inventive performance (OECD, 2009), meaning that patents can be used as a proxy for innovative activity. Second, patent data is broadly available at low cost. As patents are a legal document and assessed by patent examiners, the data in these documents is reliable and maybe more important, it is consistent over time.

There are also problems with using patent data for statistics. One of the most important issues is that not all inventions are patentable and not all inventions are patented (Archibugi & Pianta, 1996). This implies that much information, particularly tacit knowledge, is not taken into account (Malecki, 2014).

Despite these disadvantages, there are good reasons why in this research patents are used. First, for this type of geographical analysis patents are very useful as they contain the address of the inventor, which can be geocoded to the actual geographical location. Second, technology classes are assigned to patents, opening up the possibility to disaggregate them and to study a vast amount of mutual relations with other data in the patents. Third, patents can be used in time-series comparison, as patent data is consistent over time and the data has been collected for more than a century (Archibugi & Pianta, 1996).

3.1.1 Nanotechnology Patents

For this research the United States Patent and Trademark Office (USPTO) database is used to gather the patent data. Boschma et al. (2011) argue that because the US is the biggest economic market it is attractive to file a patent there. An analysis with this database is likely to represent the "worldwide stock of knowledge in a particular technology sector". Another advantage of using this source is that the database is freely available to everyone online as well as distributed by Google patents.

To keep track of nanotechnologies and to facilitate interdisciplinary research, the European Patent Office (EPO) started to work with special nanotechnology "Y01N" tags. These tags can provide more reliable results than for example keyword based indicators, and are certainly more comprehensive than the traditional B82B International Patent Classification (IPC) which encompasses only a single aspect of nanotechnology (Scheu et al., 2006). In order to unify the classification system of the USPTO and EPO, the Cooperative Patent Classification (CPC) was officially introduced in 2013. The CPC is closely related to the IPC structure and the World Intellectual Property Organization (WIPO) classification (CPC, 2014). The "Y01N" tags were moved into the "B82Y" CPC class, and are thus also applied to all relevant USPTO patents. The "B82Y" class comprises 8 distinct nanotechnology fields which are listed in table 3.1.

The patents were gathered from the online USPTO database[1]. An easier method is to use the bulk downloads from Google patents, however the CPC classes are not represented in those documents. The search term "CPCL/B82Y" was used in order to gather all the granted patents in nanotechnology. Patents granted between 13 January 1976 (first nanotechnology patent) and 02 September 2014 were downloaded, resulting in a total of 39299 nanotechnology patents. The patents are processed via a ruby script and the data is stored in a SQL-database. 7 patents returned errors as they contained no CPC classes or had no inventor locations present. As these are necessary for the analysis these 7 patents are omitted from the database.

3.1.2 Control Patents

For several parts of the analysis a control group of patents is used. The main objective of utilizing these patents is to control for already existing spatial concentration of innovation. The control patents are downloaded in the same way as the nanotechnology patents but they are randomly selected from the whole USPTO database in the same time-frame as the nanotechnology patents. 47044 patents were downloaded and 46950 of them were stored in the SQL database.

[1] `http://patft.uspto.gov`

Class	Description	Patents
B82Y5	Nanobiotechnology or nano-medicine	3097
B82Y10	Nanotechnology for information processing, storage and transmission	14801
B82Y15	Nanotechnology for interacting, sensing and actuating	1781
B82Y20	Nanotechnology for optics	6108
B82Y25	Nanomagnetism	1909
B82Y30	Nanotechnology for materials and surface science	8702
B82Y35	Methods or apparatus for measurement or analysis of nanostructures	1580
B82Y40	Manufacture or treatment of nanostructures	395
Total		**38373** [a]

Table 3.1: Nanotechnology fields in the CPC B82Y class

[a] As can be seen the total number of patents in the nanotechnology fields differs from the total number of patents in the database (39292). This difference can be attributed to the fact that some number of patents are assigned to the general "B82Y" class (849) and some to the "B82Y99/00" class, representing "Subject matter not provided for in other groups of this subclass" (69).

3.2 Location and Geocoding

One of the main aspects of this research is the analysis of spatial clustering of innovation. As said, an advantage of using patents in this case is that location data is contained in each document. Two kinds of addresses are available; applicant address and inventor address. In most studies the inventor address is used as the location of the actual invention, because the applicant address often represents the firm's headquarters which certainly does not have to correspond with the location of the invention (Deyle & Grupp, 2005).

The inventor address is listed as city and country of residence (e.g. "Paris, FR"). In order to apply spatial statistics to this data, a GPS location is required. The GPS data is acquired by geocoding the addresses with the Google Maps API[2]. This method basically uses the Google Maps engine to search the address in its database and returns the actual GPS coordinates. As Google Maps is considered a reliable tool for searching addresses it presents the researcher with little errors in the returned data and this way the manual cleaning of data is reduced to a minimum. An automated ruby script is written to retrieve the data via the API and store it in the SQL database. In nanotechnology 165 (1.7%) inventor locations returned errors, although most of these errors were in countries that are not part of the analysis (e.g. Korea, Taiwan). 17 locations were manually added to the database. For the control patents these numbers were 114 and 22 respectively.

As many patents contain more than one inventor this must be accounted for. While some

[2] https://developers.google.com/maps/

studies use only the location of the first inventor (e.g. Acs et al. (2002); Ó Huallacháin & Lee (2011)), the choice is made to apply fractional counting of the locations (e.g. if a patents has two inventors, their two locations are both assigned a factor of 0.5), following Leydesdorff & Bornmann (2012).

3.3 Data Descriptives

In order to get a better understanding of the spatial dynamics of nanotechnology, some descriptive statistics are presented in this section. In figure 3.1 the total amount of patents per year is displayed. We can see a fast exponential growth of nanotechnology patents, confirming the expected life-cycle pattern of the introduction phase in the period before 1990 and the growth phase after this. Although the effect of an overall increase in patenting activity in recent years probably also plays a role in this (see Hall, 2005), controlling for this would still show the strong growth of patenting in nanotechnology. In this research the application date of the granted patents will be used as they correspond with the actual moment when the invention is made. The drop in patents beginning from 2005 can be partly explained by the delay that is present in the application date and the date granted. However, it is realistic to assume that patent activity growth in itself also halts from 2005 onwards, as in 9 years all patents should be granted.

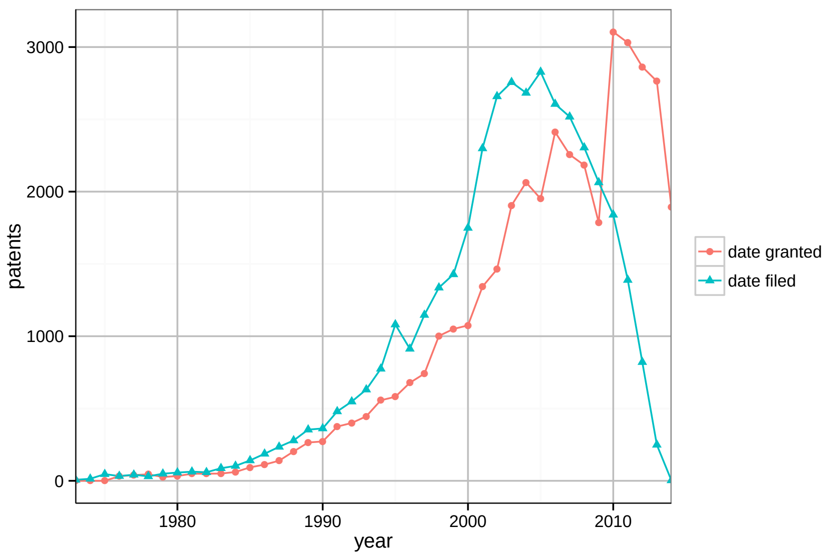


Figure 3.1: amount of patents for date granted and application date.

The patenting patterns for the different nanotechnology fields are shown in figure 3.2. The three biggest classes, B82Y10/00, B82Y20/00 and B82Y30/00 show roughly the same growth patterns. B82Y5/00, representing nanomedicine, experienced a strong growth in the mid '90s but the dropped and stabilized later on. The other technology fields grew steady over time.

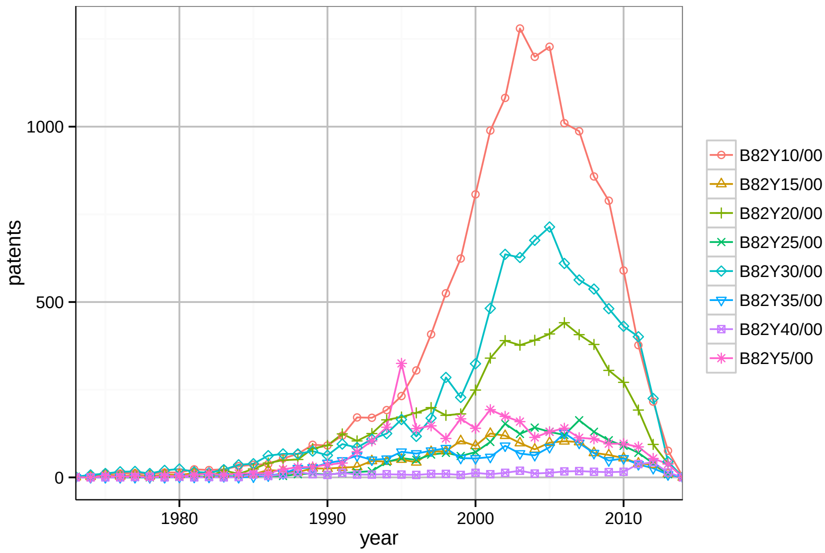
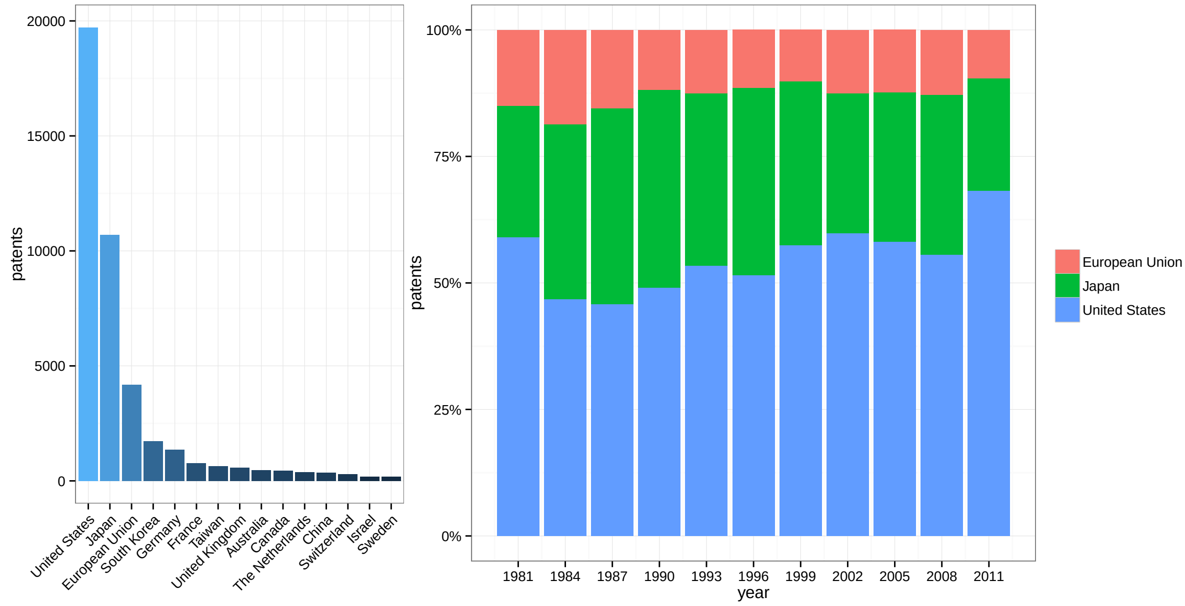


Figure 3.2: patents per nanotechnology field

It is also interesting to see how the distribution between countries has developed in nanotechnology. From figure 3.3a can be deduced that the US is by far the biggest producer of patents, although by using the USPTO database this number could be biased to the home market. Japan, the EU and Korea are after this the biggest markets. Figure 3.3b points out that the US increases in importance, while in comparison Japan and the EU both relatively decrease in patenting activity. Because of the fact that for analyzing spatial clustering we need a sufficient amount of patents[3], the rest of the thesis will be confined to three geographical areas: the US, Japan and Europe.

Before analyzing clustering patterns quantitatively, it is also useful to plot the geographical maps of the US, Japan and (part of) the EU to visualize the changing spatial patterns of the rise of nanotechnology. In figure 3.4 this transformation of the geographical landscape is illustrated by mapping the patents in the United States. The first thing that is immediately visible is that innovation activity in nanotechnology is considerably spatially concentrated. Patenting activity is concentrated in the urban areas of the US, as would be expected. Secondly, a shift in locations and their patent activity is visible. For example, while from 1993 to 1996 Denver is a very big cluster compared to the rest, in 2005-2008 San Francisco (and Silicon Valley) has taken over the lead. All other relevant geographical maps are displayed in appendix A.

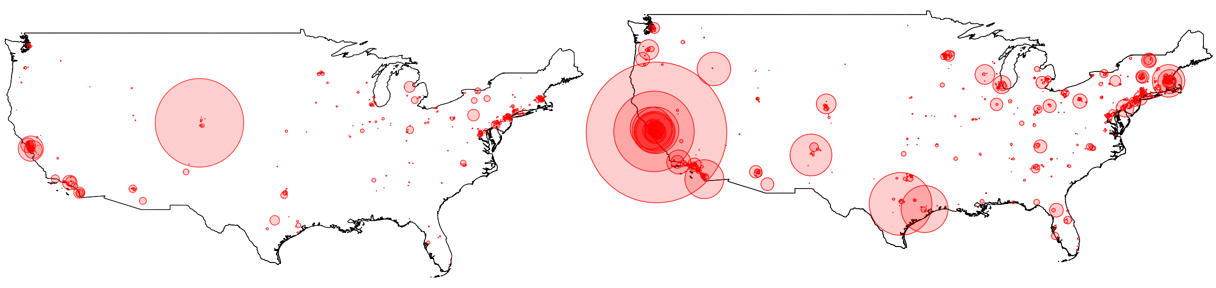
[3]This is necessary as otherwise some technology fields would have no patents in especially earlier time frames, and no spatial statistics would be possible. This would make a proper comparison difficult



(a)

(b)

Figure 3.3: (a) top 15 nanotechnology patent contributors (b) proportional representation of patents for the three major contributors in three year intervals.



(a) 1993-1996

(b) 2005-2008

Figure 3.4: geographical map of nanotechnology in the US. (a) 1993-1996, (b) 2005-2008

CHAPTER 4

Part I: Spatial Clustering of Innovation and Knowledge Base Diversity

In this chapter hypotheses 3,4 and 5 are tested. In hypothesis 3 it is expected that the number of locations increases during the early phases of the nanotechnology life-cycle. Then in hypothesis 4 we expect that the clustering of innovation decreases from the early phases on. In hypothesis 5 we expect to see a negative relation between the knowledge base diversity of nanotechnology and the spatial clustering of innovation.

4.1 Methods

4.1.1 Measuring Spatial Clustering

Measuring spatial clustering has always been of interest for economic geographers. In other fields such as ecology and epidemiology measurements of spatial clustering were already used from the 70's onward, however the analysis of economic geography proved a bit more complicated as economic activity by nature is already geographically concentrated. In the early 1990's, the Gini-index was a popular method to provide a quantification of geographical concentration of industries (e.g. Krugman, 1991a; Audretsch & Feldman, 1996a). However, this method is sensitive to the type of industry as the industry will be regarded as localized when employment is concentrated in a small number of locations (Maurel & Sédillot, 1999). Duranton & Overman (2005) put it nicely: "To understand the distinction between localization and industrial concentration, note that in an industry with no tendency for clustering, the location patterns of the plants are determined by purely idiosyncratic factors." Therefore, Ellison et al. (1997) came up with a different geographical measure which is generally referred to as the Ellison-Glaeser index (EG-index). The null hypothesis here is one of spatial randomness on both the industrial concentration and overall agglomeration of manufacturing (Duranton & Overman, 2005). Although controlling for industrial concentration was a big step forward, the method still required aggregating individual locations to regions, states or countries, which can cause statistical biases.

Duranton & Overman (2005) proposed five criteria which a localization measure has to comply with. The measure: (i) is comparable across industries; (ii) controls for the overall agglomeration of manufacturing; (iii) controls for industrial concentration; (iv) is unbiased with respect to scale and aggregation. The test should also (v) give an indication of the significance of the results. They developed the K-density function (from now on Kd-function), which satisfies these five criteria. This method is based on the k-function developed by Ripley (1976), which is a so-called distance-based measurement. This entails calculating the euclidean distance between (geographical) points resulting in a 'localization index' for each distance in the chosen range. However, the k-function cannot control for industry concentration and does not allow to compare the concentration of a sector with that of the industry (Marcon & Puech, 2010). Duranton & Overman's Kd-function has the advantages of both the EG-index and the distance-based K-function. In short[1], the kernel density of bilateral distances is estimated. If there are n points in space, the Euclidian distance between these points d_{ij} this estimation at any point d is:

$$\widehat{K}(d) = \frac{1}{n(n-1)h} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} f\left(\frac{d-d_{ij}}{h}\right) \tag{4.1}$$

In this equation a Gaussian function (representing the normal distribution) is used for f and Silverman's 'rule of thumb' is used for the bandwidth h. Distances d are within a specific chosen interval $[0,u]$.

Using equation 4.1 on its own does not provide much useful information as it is not compared to the industrial concentration, in our case the overall spatial concentration of patents. Therefore, confidence intervals can be calculated. As Duranton & Overman (2005) point out the best way to do this is by Monte Carlo simulations using n points from the 'overall population'. This population is in this case the database with randomly downloaded patents. In one simulation the bilateral distances between the random sample of points are calculated and the densities are computed. This is repeated a recommended 1000 times. The global confidence bands are, at a significance level of 5%, then defined as the values that were hit by 5% of the simulations. The industry or sector is then considered significantly localized when one or more values of its K-densities exceeds the upper global confidence interval and significantly dispersed when the values are lower than the lower confidence interval. The technical details of the Kd-function are further explained in appendix B. In this thesis a comparable clustering index is proposed based on the localization index of Duranton & Overman, which basically integrates the values above the upper confindence interval. The

[1]For a detailed explanation of the Kd-function see Duranton & Overman (2005)

equation for calculating the comparable clustering index for technology field (or sector) T is as follows:

$$CI_T = \sum_{d=0}^{d=u} \widehat{K_T}(d) - \overline{\overline{K_T}}(d) \tag{4.2}$$

where $\overline{\overline{K_T}}$ is the upper confidence band. $[0,u]$ is the chosen distance interval. Important is that only those distances d are taken into account where $\widehat{K_T} > \overline{\overline{K_T}}$. In words this equation sums the area above the upper global confidence interval and the returned value is the clustering index.

The calculation is split into two parts. First, all nanotechnology patents from the "B82Y" class are combined to analyze the developments in nanotechnology in general. Secondly, the different technology fields in the nanotechnology sector (see table 3.1), are treated separately to see if the posed hypotheses also apply on a more detailed level. In this case 7 fields are taken into account, the "B82Y40/00" class is discarded as it contains too little patents to apply the Kd-function[2]. Only patents from 1990-2011 are analyzed, as before 1990 patent activity was too low to provide good results for the Kd-function[3]. Furthermore, after 2011 patenting declines sharply as we apply the application date of the patents, while the database contains granted patents. A three-year interval is chosen for the analysis. By some initial empirical testing, this interval proved to be the minimum for the Kd-function to work with this data-set.

The distance interval $[0,u]$ must be defined because if all distances are taken into account, the K-densities would sum to one. This is because if the technology class or sector is localized in short distances, it will automatically be more dispersed at greater distances. Admittedly, the process of choosing this interval is rather arbitrarily. Duranton & Overman (2005) advise to take the median of all pairwise distances. However, for example in the case of the US this would result in a very large distance interval, as innovative activity both concentrates on east and west coast. Therefore, we adapt the procedure of Ellison et al. (2010), of initially testing multiple ranges. Their main threshold lies at about 400 km for the US. In this thesis the main results are based on a range of $[0,500]$ km.

The geocoded locations of the patents are exported from the SQL database via a Ruby script to the statistical software *R*. Then the GPS locations have to be constrained by a polygon representing the geographical area they are located in. This is done using the GIS

[2]Note that choosing a bigger time interval the sufficient amount of patents could be acquired. However, then the further conclusions about the evolution of the fields would have less power as the data would be aggregated to a greater level. Choosing the interval is thus a trade-off between reliability and usefulness of the results

[3]As each of the three geographical areas (U.S., Japan and Europe) is analyzed separately, this is especially an issue in Europe where patent activity in general is lower than the other two

shapefile maps from Eurostat[4]. Three separate GIS shapefiles from the three geographical areas that are analyzed are then created. Note that in Europe not all countries are taken into consideration. The first criterion here is that the country should have a sufficient amount of nanotechnology patents (>100). Second, the United-Kingdom, Sweden and Norway are excluded, as a sea separates them from mainland Europe. This could cause a spatial discontinuity bias with distance based measures, because then not real geographical distances would be measured (Arbia et al., 2007). For presentation purposes, the distances in decimal degrees (from the GPS coordinates) are converted to kilometers by multiplying these with 111.32[5]

The *R* package *Spatstat* provides the basic tools for doing the spatial statistics, while the package *DBMSS* extends this library with several distance-based functions, including the Kd-function.

4.1.2 Measuring Knowledge Base Diversity

The main relationship that is tested in this part of the thesis is the influence of knowledge base diversity on spatial clustering. We check not only if these two concepts correlate over time, but also try to see if a causal relationship is present. This is done by applying a lag to the independent variables to find out whether diversity precedes the development of spatial clustering, like in Neffke et al. (2011).

The data used for calculating knowledge base diversity of nanotechnology in general and as well the technology fields within nanotech, are the CPC classes assigned to the patents. Besides the assigned nanotechnology class, each patent is also assigned other classes, which are in many different fields. The aggregated result of these classes are used in this thesis as a proxy for the knowledge base in a particular field, like other contemporary literature such as Rigby (2013) and Leydesdorff et al. (2014).

The diversity of the knowledge base can be measured in many different ways. Because it is difficult to assess whether a measure is giving the proper results, two alternatives are used. The first is based on the work of Stirling (2007) who built a framework for analyzing diversity in science and technology where all three components of diversity (variety, balance and disparity) are included. He criticizes previous non-parametric diversity measurements as often they include only a subset of these three components. The diversity measure is in this thesis referred to as the Rao-Stirling[6] diversity. The formula for the Rao-Stirling diversity is:

[4]`http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco_Geographical_information_maps/popups/references/administrative_units_statistical_units_1`

[5]based on a semi-major axis of 6,378,160 m (`http://www.ga.gov.au/earth-monitoring/geodesy/geodetic-datums/historical-datums-of-australia/australian-geodetic-datum-agd.html`. $\frac{\pi d}{360}$ is then applied.

[6]as Stirling's formula is about the same as in the work of Rao (1982), while Stirling was unaware of this previous work.

$$RAO_T = \sum_{ij(i \neq j)} d_{ij} \cdot p_i \cdot p_j \tag{4.3}$$

Where d_{ij} is a disparity measure and p_i and p_j are the proportional representations of elements i and j in technology field T, measuring their balance. As can be seen, variety is incorporated by summing the proportional pairwise disparities.

The diversity is calculated by first aggregating all patents in a technological class, for example when analyzing the whole nanotechnology sector we would take all patents in the B82Y class. Then all 4-digit CPC classes that are assigned to the patents are counted. If for example the class "G08B" has a count of 300 and the "G08C" class 600 and the total number of classes is 40,000, the proportional representations of them (p_i and p_j) are 0.0075 and 0.015 respectively. The disparity between two classes is more difficult to determine. In this case, the data from Leydesdorff & Rafols (2011)[7] is used, who have constructed a patent citation co-occurrence matrix between patent classes[8]. From this data, the cosine-distance between the different classes is calculated that resembles the disparity between them. The sum across the half-matrix (as $i \neq j$) of all classes is taken, which then gives the final diversity D.

The second variable that is used is the entropy measure adopted from Frenken et al. (2007). I use this measure as it does not rely on the disparity values that are calculated in the Rao-Stirling diversity measure, so that it functions as an additional validation. Strictly, the entropy measure is not measuring diversity but only accounts for the variety and the balance of the system. Frenken et al. (2007) distinct unrelated and related variety in their work which are also both calculated in this thesis. For the calculation we take P as the proportional representation of a three-digit CPC class in technology field T and $g = 1....G$ in which G is the total amount of three-digit classes. The three-digit classes are composed of four-digit classes i , its proportional element represented by p. Unrelated variety is then calculated as:

$$UV_T = \sum_{g=1}^{G} P_g log_2 \left(\frac{1}{P_g} \right) \tag{4.4}$$

And related variety as:

$$RV_T = \sum_{g=1}^{G} P_g \sum_{i \in T_g} \frac{p_i}{P_g} log_2 \left(\frac{1}{p_i / P_g} \right) \tag{4.5}$$

[7]This data is publicly available at `http://www.leydesdorff.net/ipcmaps/`

[8]The citation matrix for calculated for IPC classes, while in this study CPC classes are used. Using CPC is considered more accurately in this thesis as they have been recently updated with assigning the patents to the B82Y class. While linking the Leydesdorff & Rafols data with CPC classes gives rise to some issues, as some CPC classes do not exist in IPC, the great majority of classes are the same in IPC and CPC.

Unrelated variety thus measures the variety in the three-digit patent classes, while related variety captures also the variety within the four-digit classes.

4.2 Results

4.2.1 Number of Locations

Testing hypothesis 3, in which I expected geographical locations to rise along the cycle, is relatively simple, as we only have to calculate the entry and exist patterns in patent locations. These patterns are plotted in figure 4.1. The number of locations are calculated by summing all the unique locations per geographical area in the three-year interval per country. The entry of locations is determined by taking the new locations that were not present in the previous three years, while exit is acquired by taking the locations that were present in the previous three years and not in the three years after this.

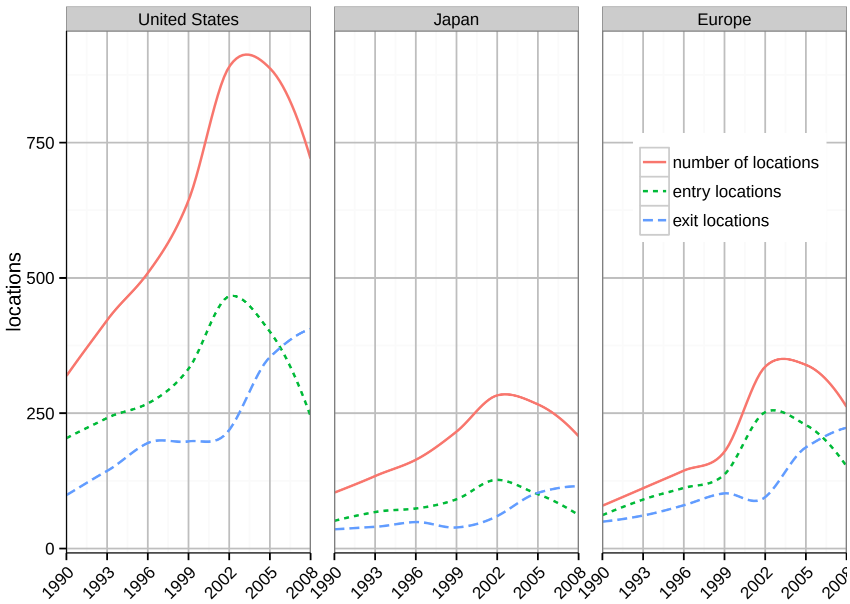


Figure 4.1: The number of unique patent locations per geographical area and the entry and exit of locations per geographical area using a three year interval

We can immediately see that development of the number of locations show the same pattern in each geographical area. In the early phase of the nanotechnology life-cycle, the number of locations rises sharply, after which the growth suddenly stops and even drops 2002-2005 onward. This seems to largely confirm what was expected in hypothesis 3. In the early period entry of locations always exceeded exit of locations, which obviously results in an increase of the unique locations. From 2002 onward the drop in locations is not only fueled by a sharp decrease in entry but also by a very apparent rise in the exit of locations. This could mean that some locations are not doing well or that firms and institutions prefer already existing locations. However, a more apparent explanation may be the drop in overall patent activity that we saw in figure 3.1. The decreasing patent activity will logically reduce the amount of small patent locations, if all other variables remain the same.

4.2.2 Clustering of Nanotechnology and Knowledge Base Diversity

In hypothesis 4 and 5 I expect the spatial clustering of innovation to decrease in time and that this is influenced by an increase in knowledge base diversity. To test these hypotheses, we first calculate the Kd-function for nanotechnology in general (the whole B82Y class) in the three-year interval per geographical area. Some results of the Kd-function in equation 4.1 are displayed in figure 4.2. Even without defining a clustering index here, valuable information can be derived from these graphs. 3 of the 7 intervals are displayed here for the three geographical areas. We see that from 1990-1993, nanotechnology is significantly localized in all areas in comparison to the control group, especially on the short distances. We see that in the later periods this localization decreases, and Europe even shows no significant clustering anymore. There are also interesting differences to note between the three geographical areas. Japan is very spatially concentrated on the short distance, what can be explained by the Tokyo urban area and the Kyoto-Osaka-Kobe cluster which comprise the majority of Japan's patent production. In Europe we see not much localization on the short distances, which can be explained by the fact that the population in general is less concentrated in big cities.

For each time period the comparable clustering index is calculated by using equation 4.2. In figure 4.3 we see that the clustering index shows the same decreasing pattern for each of the three locations, with Japan illustrating values high above the US and Europe, indicating it is indeed spatially concentrated to a great extend in nanotechnology.

Before we do any statistical analysis, it is useful to look at the graphs that visualize diversity and the variables it might influence. In figure 4.4 we can see clearly that the diversity measures precede the number of locations (NLOC) in time. Interesting is that the clustering index (CI) has the same pattern in each geographical area, but that for Japan and Europe the pattern lags behind the US. Because the diversity measures used are globally calculated, instead of per country, a lag is introduced to the diversity measures, similar to the study of Neffke et al. (2011). This calibration is done by examining the graphs and by manually adjusting the lag in each variable and per geographical area to see where the correlation is highest.[9] The applied lag is 0 years for the US, 3 years for Japan and 6 years for Europe.

[9] The possibility of introducing a similar lag for each geographical area was also tested. Although this might be more consistent in the analysis itself, as well as producing good results, it would be odd to introduce a lag in the US while the diversity measure is then actually lagging behind the clustering index. As the clustering index behaves differently in each area, the choice was made to calibrate the lag of the diversity measures for each area independently. This way also conclusions can be drawn about the 'repsonse time' of clustering on global diversity per geographical area

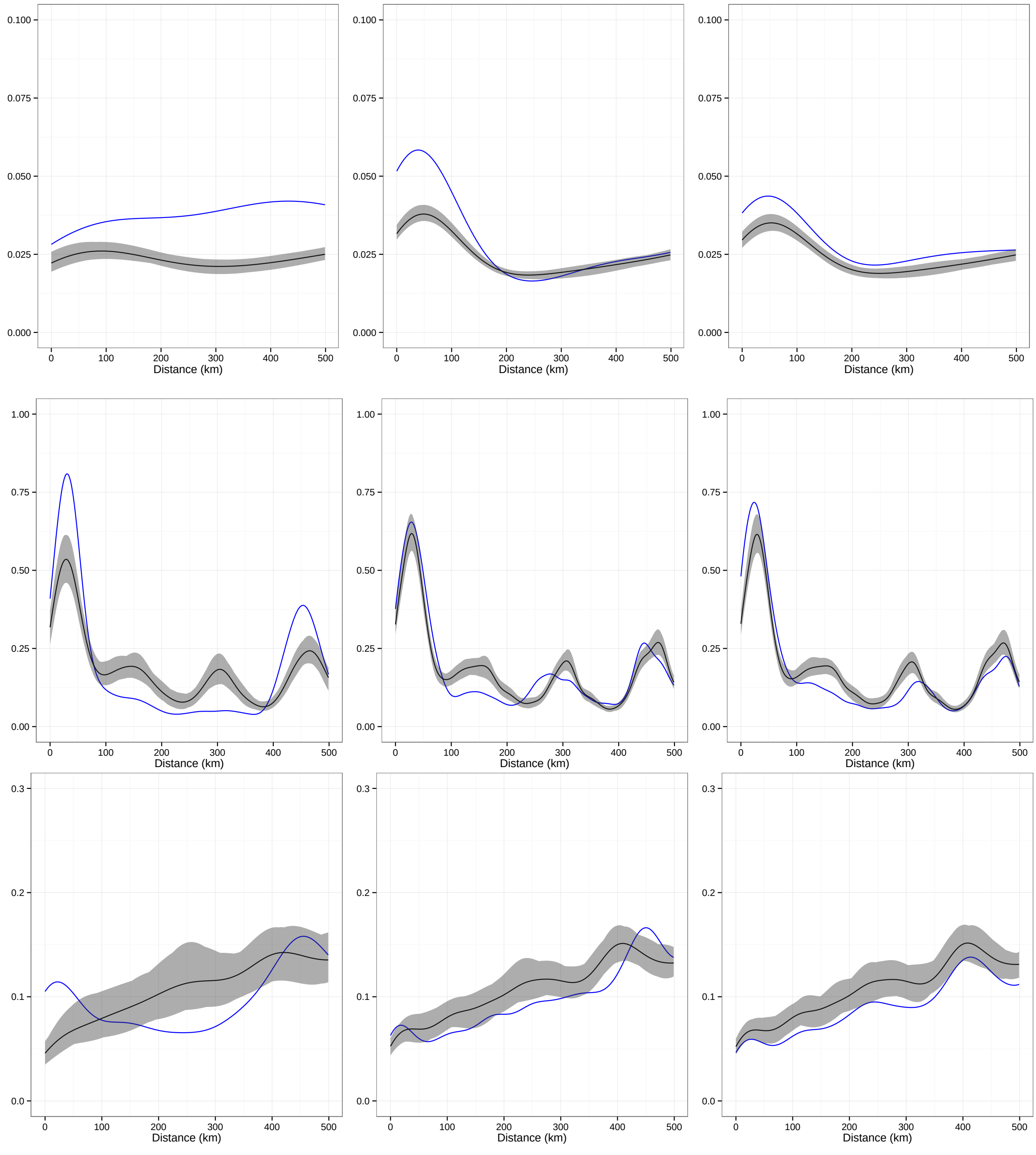


Figure 4.2: Kd-function for patents in B82Y nanotechnology class for the years (from left to right) 1990-1993, 1999-2002 and 2008-2011. From top to bottom: US, Japan and Europe. The grey area displays the confidence interval generated by the Monte Carlo simulations on the control patents

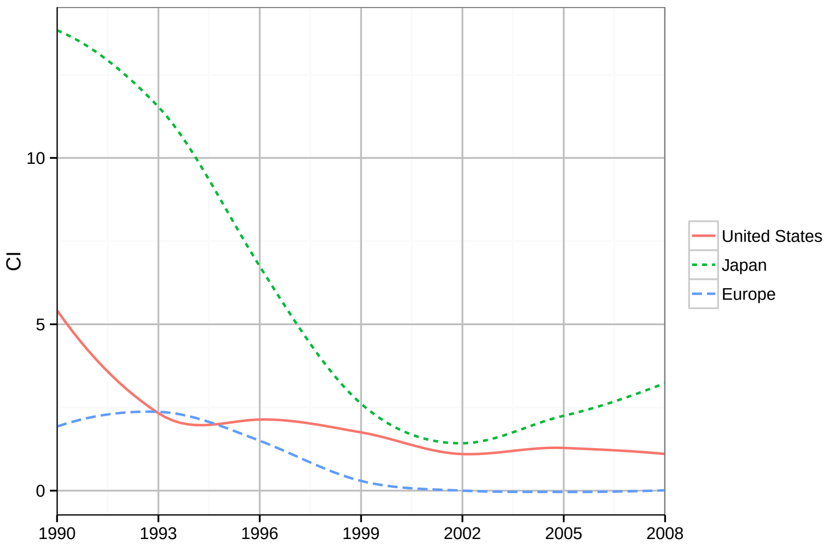


Figure 4.3: The clustering index (CI) over time for the three geographical areas

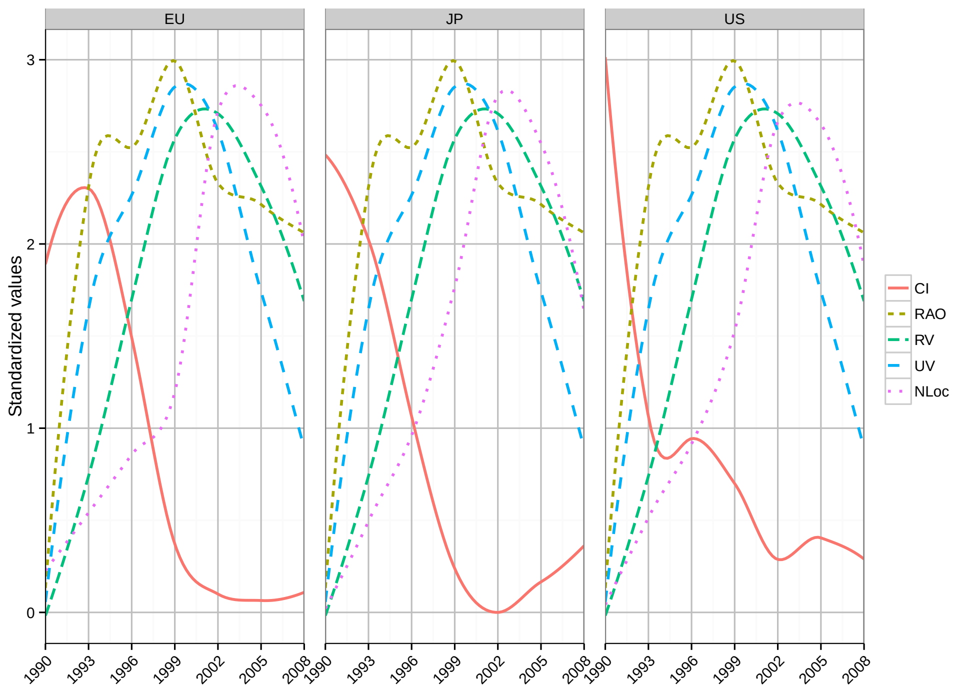


Figure 4.4: The clustering index (CI), Rao-Stirling diversity (RAO), Related Variety (RV), Unrelated Variety (UV) and number of locations (NLoc) over time in each geographical area

Because the diversity measures are equal for each of the three geographical locations, as they represent the global knowledge base, and the clustering indexes are calculated separately, the dependent and independent variables are not properly comparable. Therefore all variables used are standardized per location by the following equation:

$$z = \frac{x - \mu_a}{\sigma_a} + min(a) \tag{4.6}$$

where a is the data within the three geographical areas.

If we take a look at the Pearson correlations (table 4.1), it is apparent that all variables are highly correlated. The mutual correlations of the diversity indices point out that they all measure the same pattern in knowledge base diversity. It is also interesting to see that the number of locations, assignees and patents almost perfectly correlate with each other. The relation assignees and number of patents is fairly obvious, because when more firms enter the market the more patents are likely generated. That the number of locations correlates so well with this is less trivial, however is explainable by the fact that when overall patent activity increases the chance that inventors file patents in new locations increases as well. Apperently, the distributions of the number of patents and the number of geographical locations have a perfect linear relationship, as the correlation is almost one.

Table 4.1: Pearson-correlation matrix. NLoc=number of locations, NAsn=number of assignees, NPat=number of patents. Latter three variables are calculated per geographical location

	CI	RAO	RV	UV	NLoc	NAsn	NPat
CI	1.00						
RAO	-0.89	1.00					
RV	-0.83	0.74	1.00				
UV	-0.79	0.89	0.84	1.00			
NLoc	-0.84	0.68	0.92	0.73	1.00		
NAsn	-0.85	0.71	0.94	0.77	0.99	1.00	
NPat	-0.84	0.69	0.92	0.75	0.99	0.99	1.00

A simple OLS regression is performed to analyze the results (table 4.2). Due to the colinearity among the independent variables 4 models are created[10]. Adding the number of assignees and number of patents to any model is redundant in this case, as they correlate almost perfectly with the number of locations. The results from model 1,2 and 3 show that rao-stirling, related variety and unrelated variety all three on their own have a significant relation with the clustering index with a high explained variance. This thus implies that they can all

[10]This colinearity was tested by combining the variables into one model and then calculating the VIF (Variance Inflation Factor). This resulted in high values for the VIF, leading to the conclusion that it is problematic to combine the variables in one model. Furthermore, theoretically it would also be odd to include the number of locations in the same model as diversity, as diversity likely both influences the clustering index as well as the number of locations

be used to predict spatial clustering of innovation. However, the rao-stirling diversity fits best, as the R^2 is higher than the related and unrelated variety measures. Also important is to view these regression model in combination with the actual temporal patterns of the variables, graphed in figure 4.4. Especially in the case of the US, we can see that related variety and unrelated variety actually lag behind the clustering index a little bit, while the rao-stirling diversity matches the inverse pattern much better. In Japan and Europe the rao-stirling diversity also precedes the other measures in time, which can lead us to say that the disparity between technologies plays a role and may show earlier than just measuring variety.

Table 4.2: Results for nanotechnology (B82Y)

	Dependent variable:			
	CI			
	(1)	(2)	(3)	(4)
RAO	-0.892^{***}			
	(0.104)			
RV		-0.833^{***}		
		(0.127)		
UV			-0.792^{***}	
			(0.140)	
NLoc				-0.845^{***}
				(0.123)
Constant	2.846^{***}	2.311^{***}	2.309^{***}	2.158^{***}
	(0.242)	(0.241)	(0.276)	(0.212)
Observations	21	21	21	21
R^2	0.796	0.694	0.628	0.713
Adjusted R^2	0.786	0.678	0.608	0.698
Residual Std. Error (df = 19)	0.439	0.538	0.594	0.521
F Statistic (df = 1; 19)	74.337^{***}	43.160^{***}	32.074^{***}	47.315^{***}
Note:			*p<0.1; **p<0.05; ***p<0.01	

In hypothesis 5 I expected that the degree of spatial clustering is negatively related to the knowledge base diversity of nanotechnology in the first life-cycle phases. As temporal precedence can be observed for diversity on spatial clustering of innovation, even evidence is found for a causal relationship. The first remark that can be made about this evidence is that the apparent reaction of spatial clustering to an increase in (global) knowledge base diversity in nanotechnology seems to differ per geographical area. Furthermore, due to the temporal precedence we can compose the most likely causal model of the different variables that were measured (figure 4.5). As all diversity indictors preced the number of assignees, the number of patents and the number of geographical locations, a causal positive relation can be assumed. Theoretically, the most probable relation seems that an increase in technological diversity in the knowledge base causes an increase in the entry of firms, thus increasing the number of assignees. This than of course leads to an increase in the amount of patents. The number of assignees correlates almost perfectly with the number of geographical locations as well. The regression analysis shows that the number of locations also negatively influences the clustering of innovation. However, the diversity also seems to have a direct negative influence on the degree of spatial clustering of innovation, as the latter precedes the number of locations in time as well to some degree.

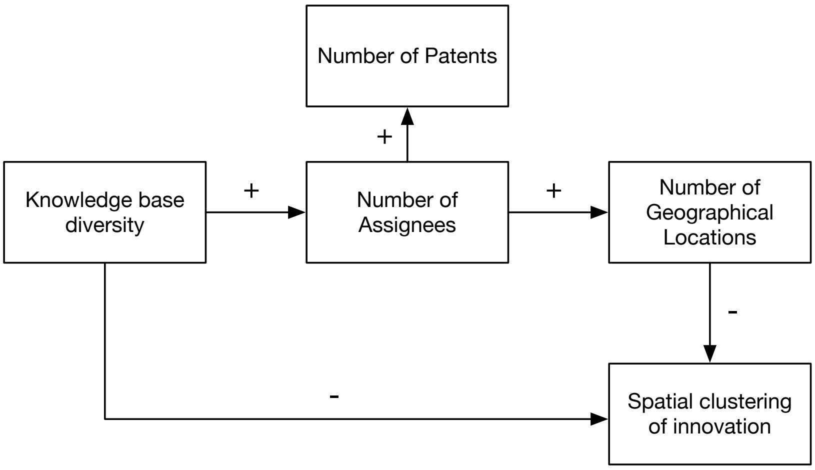


Figure 4.5: Causal scheme based on the relations found in the regression analysis

The same analysis is performed for the 7 technology fields in nanotechnology. Instead of calculating the dependent and independent variables for the B82Y class as a whole, the variables are now calculated per technology field, per year and per geographical area. The lags applied to each geographical area are the same as the previous analysis (US=0 years, Japan=3 years and Europe=6 years).

The results in table 4.3 show that the relation between diversity and spatial clustering is less obvious when analyzing the technology fields separately. The correlations are weak and the R^2 is under 0.10 in all models. Models 1 and 2 show that the related variety fits a bit better (higher R^2) in the model than the Rao-Stirling diversity, although they are both significant. Unrelated variety performas less than the latter two, with a lower R^2 (model 3). The number of locations is higly significant as well, albeit also with a low explained variance.

The difference between analyzing nanotechnology as a whole and the technology fields can be attributed to two possible explanations. First, the specific fields are possibly more sensitive to random clustering events and are thus genuinely less influenced by diversity and the number of locations. Second, the number of patents analyzed is in some particular fields quite low, which would result in more random errors in the clustering index and diversity, leading to a lower explained variance. Due to these reasons this analysis is less useful in explaining the effect of diversity than looking at the sector in general.

Table 4.3: Results for 7 nanotechnology fields

	Dependent variable:			
	CI			
	(1)	(2)	(3)	(4)
RAO	-0.254^{***}			
	(0.080)			
RV		-0.286^{***}		
		(0.080)		
UV			-0.211^{**}	
			(0.081)	
NLoc				-0.303^{***}
				(0.079)
Constant	2.031^{***}	2.153^{***}	1.997^{***}	2.169^{***}
	(0.173)	(0.186)	(0.192)	(0.181)
Observations	147	147	147	147
R^2	0.064	0.082	0.044	0.092
Adjusted R^2	0.058	0.076	0.038	0.085
Residual Std. Error (df = 145)	0.902	0.893	0.911	0.888
F Statistic (df = 1; 145)	9.979^{***}	12.949^{***}	6.724^{**}	14.610^{***}
Note:				*p<0.1; **p<0.05; ***p<0.01

CHAPTER 5

Part II: Co-localization of Technology Fields

In this chapter another aspect of spatial clustering is studied; the co-localization of the different nanotechnology fields. In hypothesis 6 I expect that the geographical co-localization of patenting activity is positively related to the degree of relatedness in the knowledge bases of technology fields.

5.1 Methods

Relatively little research has been done into the co-localization of different industries or technologies. Previously, the most used equations were the non-distance based adoptions of the EG-index and a similar measure developed by Maurel & Sédillot (1999). However, Duranton & Overman's Kd-function is also applicable for the co-localization of industries or technologies and still has the earlier discussed advantages over the non-distance based indicators. One of the more recent major study into co-localization (also often referred to as 'co-agglomeration') was by Ellison et al. (2010), who use both their EG-index and the Kd-function to test Marshall's theory of industry agglomeration. In order to use the Kd-function for co-localization for industry A and B, equation 4.1 is adjusted to the following form:

$$\hat{K}_{(A,B)}(d) = \frac{1}{P(n_A, n_B)h} \sum_{i=1}^{n_A} \sum_{\substack{j=1 \\ j \neq i}}^{n_B} f\left(\frac{d - d_{ij}}{h}\right) \tag{5.1}$$

where $P(n_A, n_B)$ is the total number of bilateral distances between *pairs of inventor locations*.

The application of the function is also different, as now we do not apply the intervals we used in chapter 4 of the thesis[1]. This time, the total amount of patents of a technology fields

[1]The aggregation of the patent locations over time brings an important advantage. The second analysis in chapter 4 of this thesis, analyzing the 7 nanotechnology fields, shows that using the three year interval causes the Kd-function to show less reliable results because of the low amount of patents. As co-localization can be studied as well without temporal information, taking the total amount of patents is preferred in this analysis

is taken from 1990-2011. Also, now not a clustering index is calculated, but a degree of co-localization. The principle is the same as in part I of this thesis; the area between the densities ($\widehat{K_A}(d)$) and the computed upper confidence interval ($\overline{\overline{K_B}}(d)$) shows this co-localization index (COI):

$$COI_{A,B} = \sum_{d=0} \widehat{K_A}(d) - \overline{\overline{K_B}}(d) \tag{5.2}$$

This equation is applied to the full matrix of the 7 by 7 technology fields, after which half of the resulting matrix is taken[2]

The knowledge base relatedness (KBR) is calculated by comparing the patent classes of a pair of nanotechnology fields. Counting all different patent classes in two nanotechnology fields results in two vectors with the length of 409 (the number of CPC classes). The cosine similarity between these two vectors X and Y is then calculated, following Leydesdorff & Rafols (2011):

$$KBR = cos(\theta) = \frac{X \cdot Y}{\| X \| \| Y \|} \tag{5.3}$$

Besides the knowledge base relatedness, we also calculate some other independent variables. The diversity measures from part I can also be applied here. It could be the case that difference of the degree of knowledge base diversity between two technology fields influences the co-locations of the two fields. The Rao-Stirling diversity (RAO), Related Variety (RV) and Unrelated Variety (UV) are calculated for each technology field the same way as in equations 4.3, 4.4 and 4.5. Now the absolute distance between the two knowledge base diversities is calculated:

$$\Delta RAO = |RAO_A - RAO_B| \tag{5.4}$$

$$\Delta RV = |RV_A - RV_B| \tag{5.5}$$

$$\Delta UV = |UV_A - UV_B| \tag{5.6}$$

5.2 Results

The co-localization index and the independent variables are calculated for each pair of the 7 distinct nanotechnology fields and for each geographical area (US, Japan and Europe). The co-localization index is standardized in the same way as the clustering index in part I. The first interesting aspect that came to light when the co-localization was calculated that Japan showed very different results than the U.S. and Europe. Japan showed high

[2] The combinations (A,B) and (B,A) have the same value, thus only half of the values minus the {(A,A),(B,B),(..,..)} diagonal are needed.

peaks on the greater distances along the Kd-function. It seems that due to the geography of Japan, its great clustering in a few cities, the co-localization should be analyzed using a lower threshold. While for the US and Europe the distance interval was maintained at [0,500 km], for Japan this was adjusted to [0,300 km].

In table 5.1 the standardized means (over the three geographical areas) of the calculated co-localization index between all the nanotechnology fields are presented. Here we see that there is quite a difference observable between the co-localization of the fields. 'Nanomechnatism' and 'Interacting, sensing and actuating' show a relatively high degree of co-localization to other fields, while 'Optics' and 'Nanobiotechnology or nano-medicine' are relatively less located near other fields.

Table 5.1: Standardized means of the co-localization index (COI) between the nanotechnology fields

	B82Y10	B82Y15	B82Y20	B82Y25	B82Y30	B82Y35	B82Y5
B82Y10: Information processing, storage and transmission	0.00						
B82Y15: Interacting, sensing and actuating	1.12	0.00					
B82Y20: Optics	0.93	0.97	0.00				
B82Y25: Nanomagnetism	0.80	2.86	0.30	0.00			
B82Y30: Materials and surface science	0.61	0.65	0.51	1.97	0.00		
B82Y35: Measurement or analysis of nanostructures	0.00	0.55	0.23	1.55	0.38	0.00	
B82Y5: Nanobiotechnology or nano-medicine	0.82	1.42	0.48	1.67	0.02	0.31	0.00
Across total:	4.28	7.57	3.42	9.15	4.14	3.02	4.72

In the correlation matrix (table 5.2) some interesting results come to light. First we see that the knowledge base relatedness between pairs of technology fields does not strongly correlate with co-localization of technology fields. Further, as expected, knowledge base relatedness correlates negatively with the degree of distance in knowledge base diversity between two technology fields. This thus implies that the more related two fields are, the lower the distance between their diversities. For example, two fields with high knowledge base diversities (resulting in a relatively low Δ RAO), are thus relatively similar. Unexpected is the negative correlation between rao-stirling diversity distance and related variety distance. As we expect they would measure more or less the same concept, a negative correlation is instead suggesting an inverse relation.

In order to see which relations are significant, a couple of regression models are composed (table 5.3). Models 1,2,3 and 4 test all the independent variables on their own. In model 5 Δ RAO and Δ RV are tested to see if they both are significant because they are significant in the other models [3].

[3] In this case the variables could be used in the same model as the VIF values were lower than 2, indicating that the collinearity of the variables was acceptable

Table 5.2: Pearson-correlation matrix

	COI	KBR	Δ RAO	Δ RV	Δ UV
COI	1.00				
KBR	-0.07	1.00			
Δ RAO	0.47	-0.41	1.00		
Δ RV	-0.41	-0.13	-0.39	1.00	
Δ UV	0.27	-0.15	0.50	0.07	1.00

The first interesting observation is that knowledge base relatedness (KBR) is not significant in model 1. This thus implies that the relatedness between knowledge bases does not really matter for the co-localization of two fields, at least not regarding relatedness between patent classes. It could well be when measuring scientific relatedness[4] the relatedness does have an impact on co-localization. However, these results are still contrary to what was expected in hypothesis 6, in which I expect that knowledge base relatedness would influence co-localization positively. Also surprising is that the difference in knowledge base diversities of the nanotechnology fields (Δ RAO) has a strong relation with co-localization in model 2. The positive relation suggest something rather odd. Fields with low knowledge base diversities tend to locate near fields with a high diversity (or vice-versa). Model 3 shows a significant inverse relation between Δ RV and co-localization. This means that fields with a high related variety of the knowledge base have a tendency to locate near other fields with a high related variety. Unrelated Variety shows the same relation as Rao-Stirling but weaker in model 4. In model 5 both Δ RAO and Δ RV are significant.

In this co-localization analysis no conclusions can be drawn like we did in chapter 4 with temporal causal precedence. However, it is likely that both the Rao-Stirling diversity and related variety influence co-localization of nanotechnology fields. The other way around might theoretically be possible, but very unlikely. The knowledge base of technology field is measured on a global level, unlikely influenced by local geographical proximity of another technology field.

[4]For example, scientific relatedness could be measured by analyzing semantic patterns in scientific journals, similar to studies by Leydesdorff & Rafols (2011) and Heimeriks & Boschma (2013)

Table 5.3: Co-localization results for U.S. and Europe

	Dependent variable:				
	COI				
	(1)	(2)	(3)	(4)	(5)
KBR	−0.066				
	(0.128)				
Δ RAO		0.465***			0.360***
		(0.113)			(0.119)
Δ RV			−0.411***		−0.271**
			(0.117)		(0.119)
Δ UV				0.268**	
				(0.123)	
Constant	1.168***	0.527***	1.552***	0.665***	0.975***
	(0.254)	(0.169)	(0.182)	(0.215)	(0.256)
Observations	63	63	63	63	63
R^2	0.004	0.217	0.169	0.072	0.279
Adjusted R^2	-0.012	0.204	0.155	0.057	0.255
Residual Std. Error	0.990 (df = 61)	0.878 (df = 61)	0.904 (df = 61)	0.956 (df = 61)	0.849 (df = 60)
F Statistic	0.271 (df = 1; 61)	16.871*** (df = 1; 61)	12.400*** (df = 1; 61)	4.716** (df = 1; 61)	11.592*** (df = 2; 60)

Note: *p<0.1; **p<0.05; ***p<0.01

CHAPTER 6

Part III: Cluster Performance, Diversification and Related Variety

In this chapter the relative performance of clusters is analyzed. In hypothesis 1 I expect that the more diversified a nanotechnology cluster is, the better it performs. Hypothesis 2 expects that performance is higher when the cluster has a higher related variety in the nanotechnology knowledge base.

6.1 Methods

In this analysis only the biggest clusters are taken into account, as the smaller the amount of patents will get, the more unreliable the calculations for diversification and related variety will be. Making a list of all clusters is difficult however, as the clusters, often urban areas, include many different smaller patent locations. All clusters thus have to be manually identified. Therefore, the clusters are identified by simply looking at the maps of overall patenting activity in each geographical area. In the end the 20 biggest nanotechnology clusters are selected, see table 6.1. For Japan only three clusters are selected as almost all patenting in nanotechnology is concentrated in these three regions. The amount of patents in a cluster is determined by taking all locations within a 70 km range from the center of the cluster.

The first objective is to identify the relative performance of nanotechnology clusters. There are many different methods that could be applied to see which clusters perform better and which worse. In studies like Boschma & Wenting (2007) and Buenstorf & Klepper (2009) the performance of firms is studied, via proxy variables such as firm age and firm hazard rate. These variables can be compared to the locations of these firms, to test if 'location matters'. In this case however, we do not have much coherent information about the firms inside the clusters, besides entry and exit patterns of patent assignees. The latter however does not

United States		Japan		Europe	
Silicon Valley	4499	Tokyo	5596	Paris	342
Boston	1411	Kyoto-Osaka	4464	Frankfurt	302
San Diego	705	Niigata	309	Eindhoven	243
Minneapolis	508			Zurich	219
Albuquerque	457			Cologne	140
New York	442			Munich	139
Chicago	442			Lyon	136
Houston	373				
Austin	377				
Philadelphia	367				

Table 6.1: Clusters per geographical area. Number represent the amount of patents in the cluster.

provide unambiguous performance results, as clusters with little entrants do not necessarily have to perform badly.

Instead, cluster performance is measured by calculating the patent output of the cluster in comparison to the overall patent activity. This is in line with Hidalgo et al. (2007), who introduce the concept of revealed comparative advantage (RCA). The RCA measures whether a country exports more of particular good than the average country. However, the measure used in this research is adjusted to only calculate the total amount of patents of a cluster in a particular time frame in relation to the global amount of patents in that time frame. Therefore, the relative performance (RP) of a cluster C is measured as follows:

$$RP_{C,u} = \frac{x(C,u)}{\sum x(u)} \tag{6.1}$$

where u is the same three-year interval used in earlier equations and x the amount of patents.

The first independent variable is the diversification of nanotechnology in a cluster. There are many different indicators that could be applied to measure diversification and they do show varied results in measuring this (Palan, 2010). Here is chosen to use the Shannon entropy measure, much in the same way variety is measured in the previous analyses. This method is chosen because it is likely to measure diversification in a more consistent way with respect to the other variables. Equation 4.4 is used in the same way for cluster C:

$$DIV_C = \sum_{g=1}^{G} P_g log_2 \left(\frac{1}{P_g} \right) \tag{6.2}$$

only now g are the 7 nanotechnology fields and P is the proportional representation of a nanotechnology field.

The other independent variable, related variety of the nanotechnology cluster, is calculated by equation 4.5, now only taking the nanotechnology patents within the specific cluster. For both the diversification of and the related variety of the clusters a lag of 3 years is applied, as the effects on patent production likely take time to show. The validity of applying this lag is corroborated by testing different lag configurations.

6.2 Results

In the same way as the previous analyses, a regression model is set-up to test the relations among the dependent and independent variables. For the relative performance (RP) of the clusters the logarithm is taken, as the distribution in patents among clusters is highly skewed. In table 6.2 the correlation matrix shows that both independent variables correlate with the relative cluster performance. Furthermore, it shows that diversification also correlates with the related variety in a cluster, thus the higher the diversification in nanotechnology, the higher the related variety will become. This makes sense from a theoretical point of view, as branching out to other nanotechnologies is likely to result in a higher variety in the knowledge base.

Table 6.2: Pearson-correlation matrix.

	log(RP)	DIV	RV
log(RP)	1.00		
DIV	0.41	1.00	
RV	0.45	0.40	1.00

In table 6.3 the results of the regression are shown. Both diversification and related variety in the clusters show a highly significant relation to the related performance of the cluster in nanotechnology. As we introduced a three year lag in the independent variables, we can also infer that not only a correlation is present, but as well a causal preceding of the events[1]. I conclude thus that both diversification and related variety positively influence the patent production of the clusters, thus confirming hypotheses 1 and 2. Although it must be noted that the explained variance is not that high (adjusted $R^2 = 0.249$), implying that there are other factors that determine a decline or a rise in cluster performance.

[1]The independent variables were also tested with no lag applied. This lead to a considerable lower R^2, meaning a less good fit in the regression

Table 6.3: Results for relative cluster performance

	Dependent variable:
	log(RP)
DIV	0.527^{***}
	(0.155)
RV	1.738^{***}
	(0.400)
Constant	-6.188^{***}
	(0.300)
Observations	146
R^2	0.259
Adjusted R^2	0.249
Residual Std. Error	1.005 (df = 143)
F Statistic	25.053^{***} (df = 2; 143)
Note:	*p<0.1; **p<0.05; ***p<0.01

CHAPTER 7

Discussion

In the first part of this thesis the influence of knowledge base diversity on the spatial clustering of innovation was studied along the early phases of the nanotechnology life-cycle. The life-cycle of nanotechnology shows the expected stylized pattern of the s-shape (Andersen, 1999); the number of patents in the nanotechnology sector first grows steadily in the 1980's and early 1990's (the introduction phase) after which it suddenly grows exponentially until about 2005 (the growth phase), after which it stabilizes. This confirms the theory of Klepper (1996) which predicts that innovative activity is highest in the early phases of the cycle. This growth of patents is caused by the entry of firms and institutions (or: patent assignees) in the market, which shows a very similar pattern to the amount of patents in all the three geographical areas studied (the US, Japan and Europe). The same growth pattern is seen for the amount of distinct geographical patent locations throughout these areas, confirming hypothesis 3. In fact, the number of patents, number of assignees and number of geographical locations turn out to be almost perfectly correlated with each other.

In hypothesis 4 I expected that while the number of locations would rise, the spatial clustering of innovation would drop throughout the early life-cycle phases. For the US, Japan and Europe a strong reduction of the calculated clustering index was observed, thus confirming this hypothesis. This is in line with studies from Audretsch & Feldman (1996b) and Dumais et al. (2002), who recored a decline in geographical concentration of industrial plants along the cycle. Dumais et al. (2002) attributed this to the emergence of new plant locations that causes the geographical concentration to fall. The regression analysis in this study showed this connection as well. However, the calculated spatial clustering index precedes the growth of locations in time, suggesting that other forces influence the propensity to clustering as well.

This research provided evidence that knowledge base diversity influences the number of geographical locations indirectly as well as having a direct influence on the spatial clustering of innovation in the early life-cycle. Knowledge base diversity has a high negative correlation with the clustering index. In Japan and Europe a notable lag between diversity and clustering was noticed, while the US spatial clustering seems to 'react' almost immediately to the increasing diversity. The difference between the geographical areas may be explained by the fact that diversity is calculated on a global level and that the US is the greatest contributer to this knowledge base, likely resulting in a shorter lag between the entry of firms and thus the spatial clustering of innovation in this area. The lag present in Japan and Europe seems to indicate temporal precedence, by which we may assume a causal relationship between knowledge base diversity and spatial clustering of innovation (hypothesis 5). Based on the results of the analysis the following causal scheme was composed (figure 7.1).

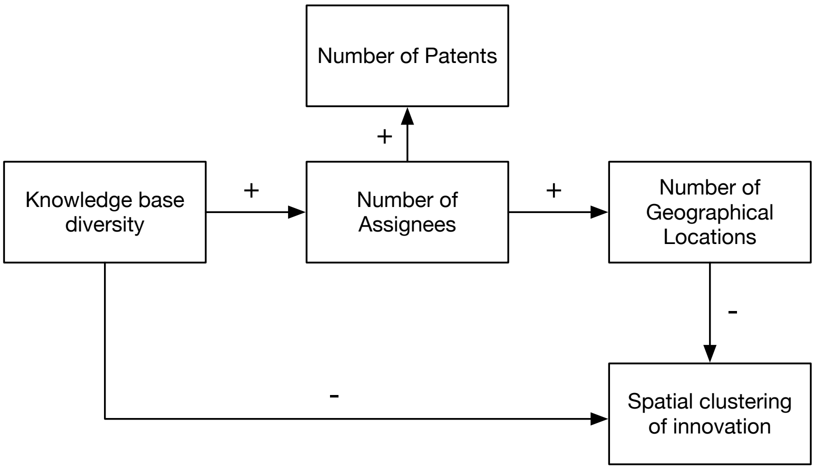


Figure 7.1: Causal scheme based on the relations found in the regression analysis (repeated figure 4.5)

It is difficult to be absolutely sure about these relations, as in theory the diversity pattern could simply co-vary with spatial clustering. However, that would mean that geography influences knowledge base characteristics, which has never been suggested in literature and is very unlikely. The evidence is certainly strengthened by the fact that the same pattern is shown in all three studied geographical areas. The resulting theory that follows from this analysis is thus that diversity in the knowledge base positively influences the entry of firms in the market and is at the same time part of an interactive system that changes the dynamics of the knowledge base in the nanotechnology sector, causing spatial clustering to become less important. Still relatively little is known about the precise workings of these knowledge dynamics. According to literature, two aspects of the knowledge base may influence spatial clustering and the need for geographical proximity. First, the increasing codification of knowledge along the cycle somehow "frees it from its geographical shackles" (Howells, 2002). Second, technological complexity with regard to the innovative process and recombinatorial difficulty decrease along the cycle (Sorenson et al., 2006), as more knowledge about the key technologies becomes known. A lower complexity decreases the need for localized learning and thus the need for geographical proximity (Malmberg & Maskell, 2010). The results from this study are perfectly in line with these theories, as spatial clustering decreases in the early phases of the cycle. It seems that the diversity of the knowledge base is connected to the development of tacit knowledge and complexity of knowledge. It remains speculating how exactly these concepts interact, but I expect that the decreasing complexity (and uncertainty) about the core technologies in nanotechnology causes the codification of knowledge to increase. This in turn opens up new possibilities for other application and research areas, leading to increasing diversity in the knowledge base. However, it could well be that diversity is not only a proxy for complexity and tacit knowledge, but that diversity itself changes the dynamics of the knowledge base and the need for geographical proximity as well. Albeit that little is empirically known about these elusive concepts, they are likely intertwined considerably.

Besides the evidence this thesis provides about knowledge base diversity influencing general patterns of clustering, the results from part II of the research also provide evidence that knowledge base diversity influences the co-location of innovative activity between various technology fields as well. This is in line with previous studies which have pointed out that co-localization of industrial plants occur between certain sectors and this effect could be attributed to MAR or Jacobs externalities (Ellison et al., 1997; Braunerhjelm & Borgman, 2006), which indicates that knowledge base characteristics play a role. The results from my analysis show that the difference in knowledge base diversity between fields correlates positively with the degree of co-localization, in other words; innovative activity in technology fields with low diversity is located near fields with high diversity. Oddly, I found no evidence for hypothesis 6, where I expected that knowledge base relatedness would affect co-localization. However, the difference in related variety (measuring the degree of relatedness of the technology classes *within* a nanotechnology field) between fields affects their co-localization negatively; fields with similar degrees of related variety thus tends be in closer proximity to each other. The positive effect of diversity and negative effect of related variety are quite surprising. The most likely explanation is that firms in technology fields with higher diversity, already further developed, diversify into younger, less developed fields which still have a low diversity[1]. These fields are not necessarily related in technology applications (knowledge base relatedness) but do have a similar degree of relatedness within the field (related variety). This would add another perspective to the geographical implications of technological relatedness (see Boschma & Frenken, 2011a; Rigby, 2013). However, this theory remains very speculative and although the results are significant, it can be coincidental in the evolution of the nanotechnology sector, given that there is no earlier empirical research that looked into these relations.

In the third part of this thesis large nanotechnology clusters have been studied. Specifically, the effects of diversification of nanotechnology fields within the cluster and related variety of the cluster on the relative performance of the nanotechnology clusters are tested. The results

[1] I expect that there is a relation younger the technology field, the lower the diversity still is. However, this is difficult to test or control for with the data available from the patent database. Qualitative data is needed to assess the maturity of the technologies and knowledge involved

are in line with contemporary literature on related variety and diversification (Frenken et al., 2007; Maggioni, 1999); both the degree of diversification into different nanotechnology fields and related variety have a positive effect on the relative cluster performance. While other studies have provided evidence for the effects on production, labor and firm entry, the results from this thesis confirm that diversification and related variety also increase innovative performance of clusters. We can also fairly certain claim that this relation is causal and not simply a correlation, as a temporal lag of 3 years on the independent variables gives the best results in the regression models. There are two possible explanations why innovative performance can increase in clusters: increasing performance of incumbent firms inside the cluster or an increase in entrants (or decrease in exit)[2]. Future research should point out which explanation fits best, or that both firm performance and entry are affected.

In general, the empirical evidence suggests that the clustering pattern of innovation in the early stages of the life-cycle is a very complex process, especially when combining these results to the contemporary literature on innovation and geography. The majority of studies in economic geography take a regional perspective and study the effects of properties of regions and its consequences for these regions, such as productivity, firm entry and innovativeness. This study showed that looking through a macro lens on spatial clustering delivers other patterns than regional focused studies would present. I think externalities and spin-off dynamics are important in explaining the path-dependent process of geographical economic evolution and provide answers why some regions are successful. While they explain local differences, they do not answer why global clustering patterns differ over time. As predicted in evolutionary economic theory (Boschma & Frenken, 2011a), geographical patterns seem to depend on the phase of the industry life cycle. Properties of the knowledge base of a sector or industry change along the cycle and this thesis provided evidence that at least one of these properties, diversity, has its effect on spatial clustering. While we know that concepts relevant to the knowledge base such as tacit knowledge (Asheim & Gertler, 2005) and complexity (Sorenson et al., 2006) influence geography, there is still little known about their effects on a global level throughout time[3]. As Boschma & Frenken (2011a) suggest, we should move beyond the 'MAR vs Jacobs' debate with regard to the agglomeration forces. I would add to this that the macro-perspective on geographical evolution, which has its effects as well on regional dynamics, should receive more attention in empirical studies within evolutionary economic geography.

[2]The reason for not analyzing the entry- and exit patters in the clusters is that the amount of firms in most clusters is not high enough to produce reliable results

[3]Noted must be that these concepts are difficult to measure, especially on a bigger scale

7.1 Limitations, reliability and validity

A limitation is the amount of data in this study. Although the nanotechnology patent database is sufficiently big for a quantitative analysis, the calculation with a distance based function requires a certain amount of patents in each time frame to give good results. As the amount of patents is low in the very early phases of the life-cycle, this time period (1980 - 1990) could not be taken into account for this study. However, as in the early 1990's nanotechnology was still in its exploratory phase, the analysis still captures the most important parts of nanotechnology development in the life-cycle. To further enhance validity, the method was adapted to use a three year time-frame, which is basically a trade-off between accuracy and the internal validity.

Secondly, the generalizability to other industrial sectors is questionable. As research has pointed out, spatial clustering can greatly differ between sectors (Duranton & Overman, 2005). For this reason, the conclusions from this research should be limited to the nanotechnology sector. Instead of including more sectors in the analysis, I choose to focus on one sector and provide a deeper analysis researching multiple aspects of spatial clustering.

Thirdly, there are some drawbacks of using the K-density function developed by Duranton & Overman (2005). Although an inherent aspect of a distance-based function, the choice of the threshold distance is a rather arbitrary. Further research should point to more sound criteria to choose this distance. Another important limitation for the applicability in other studies may be that the calculation of the clustering index itself can be problematic. In this thesis this was not so much an issue, as nanotechnology is significantly more concentrated than the control group. However, when using the Kd-function in a study that analyses multiple sectors or industries, the clustering index results in zero for sectors that are more dispersed. This way the results are then not mutually comparable as used in this study[4]. But overall, the Kd-function proved to be a valuable tool for analyzing spatial clustering in innovation sciences, as controlling for overall innovative activity concentration is implemented in the function. It is an improvement from functions that use spatial zoning, as this can give flawed results. Furthermore, for patent data it is easily applicable to multiple countries or geographical areas, as shown in this thesis. In my opinion, the Kd-function, or other developed distance-based functions provide a good modeling framework for economic geographers.

Fourth, although reliability is not so much an issue with patent data, the validity of using patents can be a problem. The main issue with patent data is the construct validity. Can patents really be used as a proxy to measure knowledge production? Although this question always remains open when using patents, this data is the only way to do this kind of quantitative analysis. One advantage of using patent data is the public availability. This increases the reproducibility, as the USPTO patent database can be accessed by anyone. Once

[4]There are other frameworks available, summarized by Marcon & Puech (2011). For example, the Kd-function and the M-function can be used as complementary tools

a patent is granted, the later changes made to it are minimal, resulting in a great reliability for the following quantitative analysis. The other data used, for example the data from Leydesdorff & Rafols (2011) and the *R* packages containing the Kd-function, are publicly available as well.

Lastly, there are some other minor treats to internal validity as well in the analysis. First, the citation data adopted from Leydesdorff & Rafols (2011) does not completely match up with the CPC classes used in this analysis. Although a little error can arise from this in calculating knowledge base diversity, the CPC classes almost perfectly fit with the IPC classes in the data. Given that about 150.000 instances of classes are linked to the patents, this error is likely to be very small. Second, the consecutive steps of calculating the area above the 5% significant confidence bands in the Kd-function and then performing a regression on these 'observations' could potentially result in a lower final significance of the regression. However, due to the fact that 1000 monte carlo simulation are run, and that a comparable clustering index is calculated[5] the error is reduced to a minimum. Thirdly, in part III of the thesis, the determination of the threshold for selecting the clusters is rather arbitrary, which could lead to a selection bias in the results. However, due to the nature of the geographical data, there is no statistical tool present to provide us with a method to determine whether a cluster has enough patents to provide a proper 'observation'. Furthermore, it must be noted that in this specific regression analysis control variables are more important than in the analysis of the global clustering patterns, as regions are more susceptible to coincidental changes and influences from their environments. The amount of 20 clusters and the fact that the variables are calculated in 7 intervals should reduce this possible error.

[5]Would we have wanted to determine if the sector is significantly localized in comparison to total industry concentration, the confidence interval of the Kd-function plays a much larger role. Now, we compare the temporal values of the clustering index with each other. A systematic error is thus much less likely as it would present itself in all calculations of the clustering index

CHAPTER 8

Conclusions

The main aim of this study was to find the influence of knowledge base diversity on the spatial clustering of innovation in the nanotechnology sector. As part of spatial clustering I also analyzed the relationship between the relatedness of knowledge bases and co-localization of nanotechnology fields. Lastly, this study also researched the performance of specific nanotechnology clusters and tested the influence of diversification and related variety. This thesis utilizes a novel approach to globally quantify spatial clustering by calculating a comparable, distance based, clustering index in order to identify evolutionary patterns. The following conclusions are drawn:

1. A clear pattern emerged regarding the development of nanotechnology over time. This pattern matches with what was expected of a sector in the early stages of its life-cycle. For all the three geographical areas analyzed (the US, Japan and Europe), both the number of firms as well as the number of geographical patent locations grew exponentially in the period 1990-2005. From there on, the number of unique firms and locations stabilizes.

2. Nanotechnology is especially geographically concentrated on the short distances (0-100km) in comparison to overall clustering of innovation, but this effect diminishes during the cycle. In Japan the clustering of innovation in nanotechnology is significantly higher than in Europe and the US.

3. The US, Japan and Europe all showed the same, fast declining, pattern of spatial clustering in the early phases of the nanotechnology life-cycle. Although the pattern is the same, Japan and Europe lag behind in time on the US. Regression analysis shows that the rise of knowledge base diversity in the early cycle is highly negatively related to spatial clustering and that this is likely a causal relation, given by the lag that Japan and Europe show. Furthermore, the development of knowledge base diversity precedes the amount of firms, patents and geographical locations in time as well. The theory I pose for the nanotechnology sector is that a higher knowledge base diversity has a

direct influence on spatial clustering, as a higher diversity goes along with less tacit knowledge and complexity in the knowledge base, causing the need for proximity to decline in the early life cycle. Furthermore, a rising knowledge base diversity seems to increase firm entry, in turn leading to a similar increase in the amount of geographical locations. This dispersion of innovative activity then results in a lower degree of spatial clustering.

4. No evidence was found for a relation between the co-localization of nanotechnology fields and the relatedness of their knowledge bases. However, a connection was found between the difference in knowledge base diversity between nanotechnology fields and their co-localization. This means that fields with lower knowledge base diversity tend to co-locate with fields that have a high knowledge base diversity. The most likely explanation is that firms in technology fields with higher diversity, already further developed, diversify into younger, less developed fields with a lower diversity. In turn, the results also showed that fields with similar related variety of their knowledge base tend to be co-located to a higher degree. Firms thus seem to diversify into fields with similar related variety. I must add that theorizing on this part remains speculative as still little is known about global patterns of co-localization of technology fields.

5. A significant positive effect of diversification of nanotechnology fields in a cluster on the relative performance of a nanotechnology cluster was found. Also, an increase in related variety of nanotechnology clearly has a significant positive influence on cluster performance. These findings confirm previous empirical research on clusters and regions, and add to this that these ’cluster variables’ also affect innovative activity in particular.

The general conclusion from this thesis is that not only regional knowledge base parameters influence spatial clustering, but that throughout the early life-cycle phases of the nanotechnology sector the changing knowledge base characteristics, and in particular diversity, influence both the propensity to spatially cluster as well as the dispersion of geographical locations on a global level.

8.1 Future research

Due to the relatively new distance-based measure used in this thesis, the computed comparable clustering index has provided us with new insights in the behavior of clustering of innovation and its relation to the knowledge base diversity. Still little research has been done in economics with distance-based spatial analysis, and I think it can supply the field with more detailed information about spatial clustering.

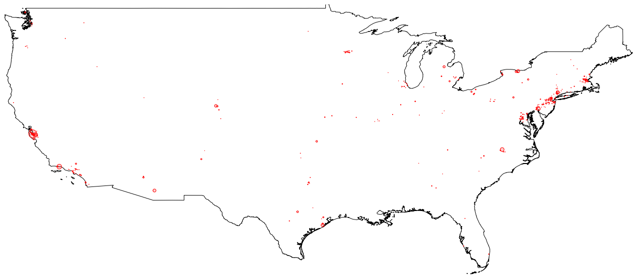
First, in this study I have only researched one specific high-tech sector. In order to be able to draw stronger conclusions regarding the influence of the knowledge base in a sector of technology, future research should include more than one sector. It would be interesting to see what the differences are between sectors and if high-tech sectors behave in another way than low-tech sectors. Furthermore, in this thesis only the beginning of the life-cycle of nanotechnology is analyzed. Other, more mature sectors, could provide a more complete picture of the other phases of the cycle.

Second, a great challenge is to discover how the diversity of the knowledge base is related to the evolution of tacit knowledge and complexity along the life-cycle. The main problem is with the measurement of these concepts, especially the amount of tacit knowledge that is involved, as it cannot be captured by the the regularly used patent data and R&D investments. The increasing utilization of qualitative data in innovation surveys could provide a solution, and it could perhaps be linked to the quantitative data we have on knowledge base diversity and spatial clustering.

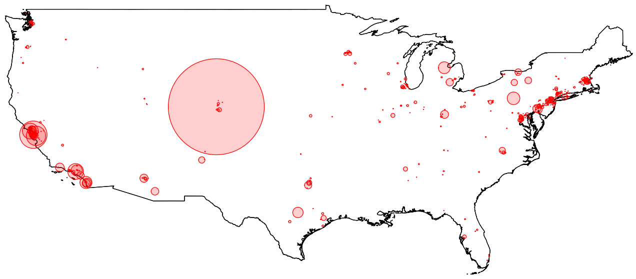
Third, on a more methodological note, it would be very promising for future research to incorporate the properties of geographical locations into the analysis of spatial clustering. It is already possible with current distance-based analyses to assign a weight to the spatial points, but as far as I know only employment numbers of industrial plants have been used as controls in this regard. One could think of incorporating data into the spatial analysis that is currently used in regional studies in which zoning is applied, such as regional labor, population or educational data, in order to compare the differences in the output of the spatial analysis. Furthermore, in the temporal regression analysis done on the clustering index more variables could be added to find more connections or to control for certain phenomena.

APPENDIX A

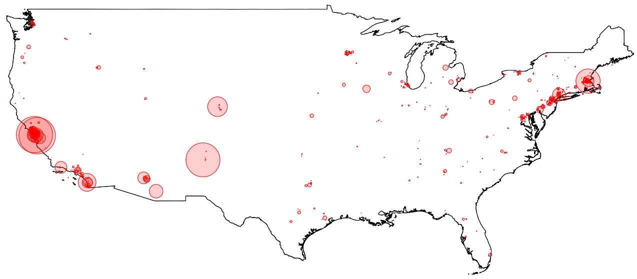
Clustering maps



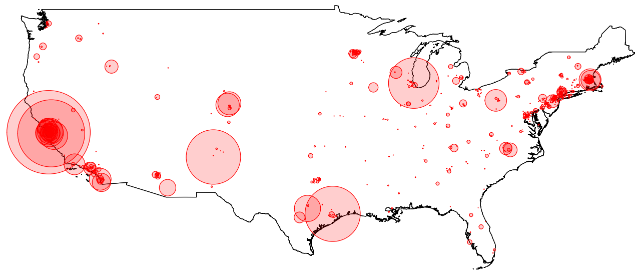
(a) 1990-1993



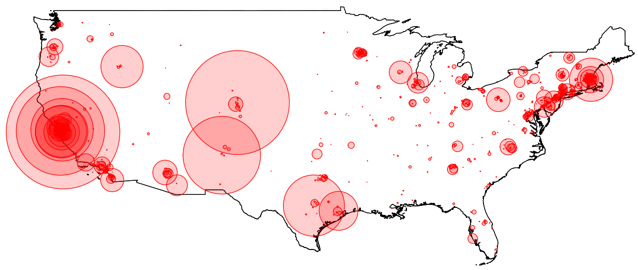
(b) 1993-1996



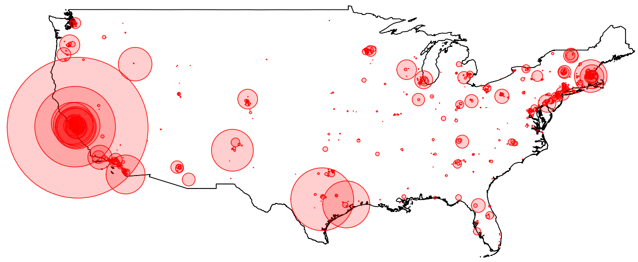
(c) 1996-1999



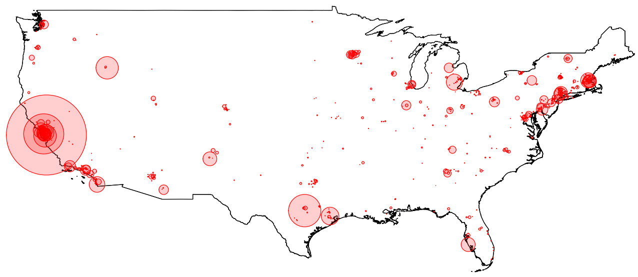
(d) 1999-2002



(e) 2002-2005



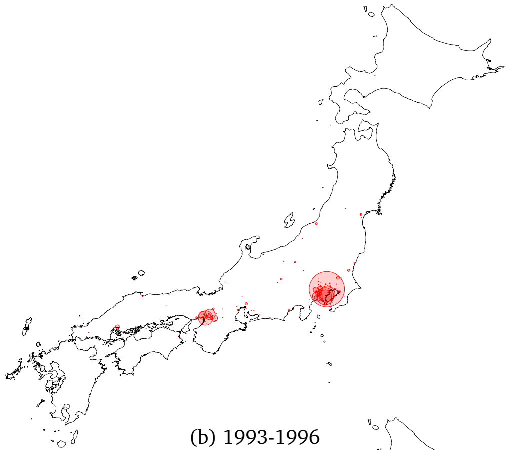
(f) 2005-2008



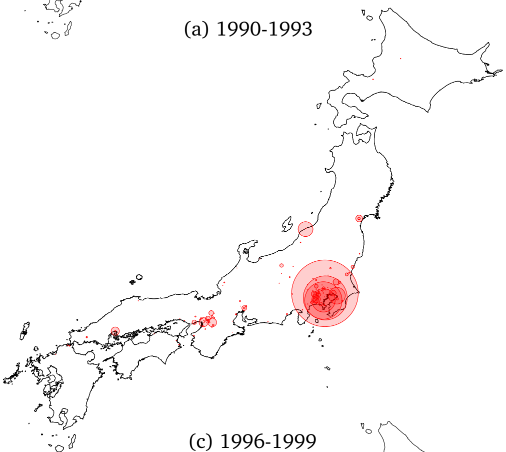
(g) 2008-2011

Figure A.1: Nanotechnology patent activity in the US over time.

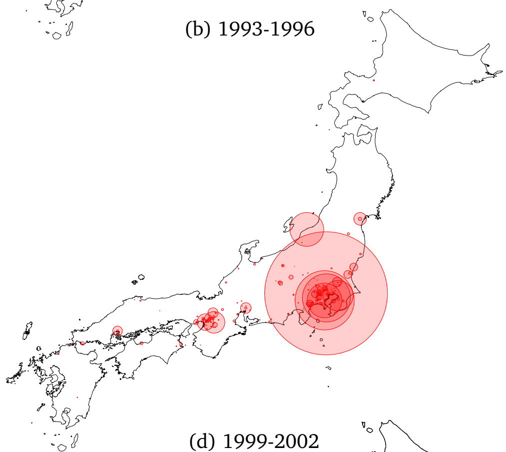
(a) 1990-1993



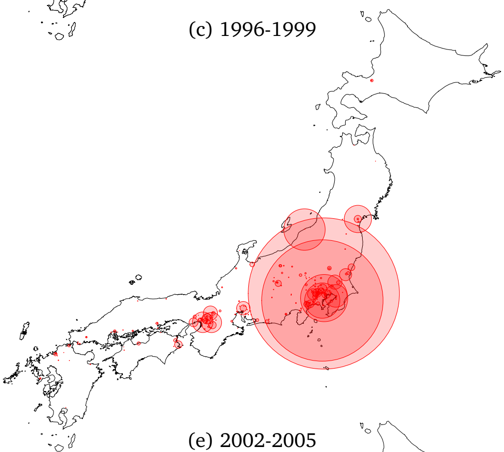
(b) 1993-1996



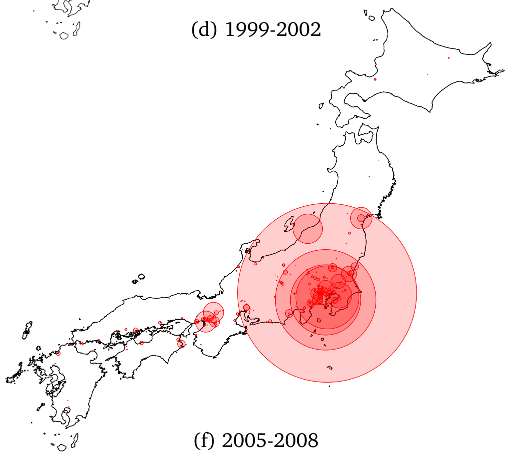
(c) 1996-1999



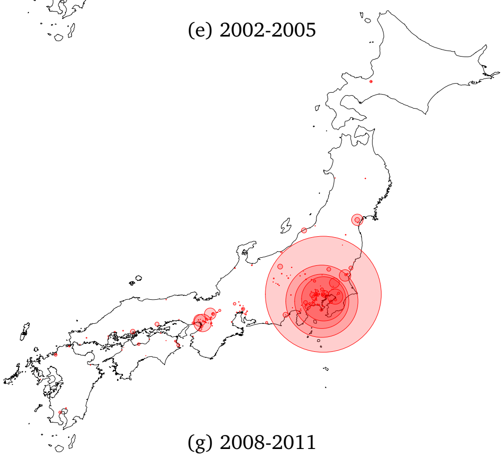
(d) 1999-2002



(e) 2002-2005

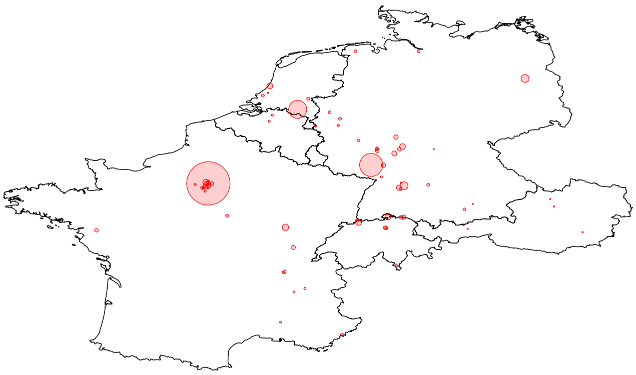


(f) 2005-2008

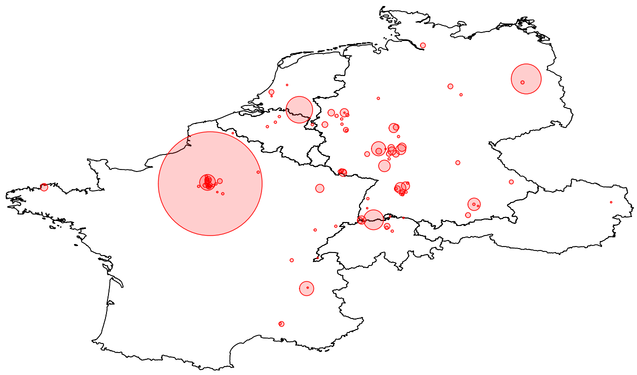


(g) 2008-2011

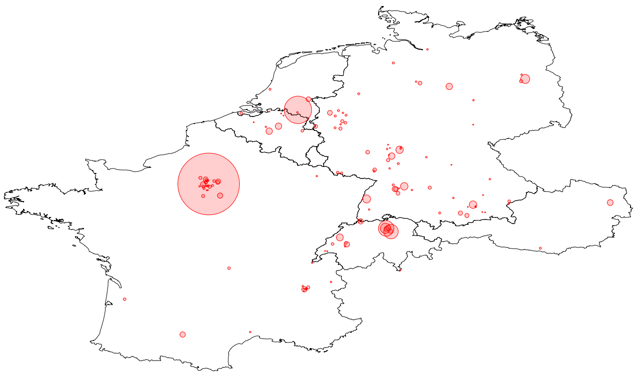
Figure A.2: Nanotechnology patent activity in Japan over time.



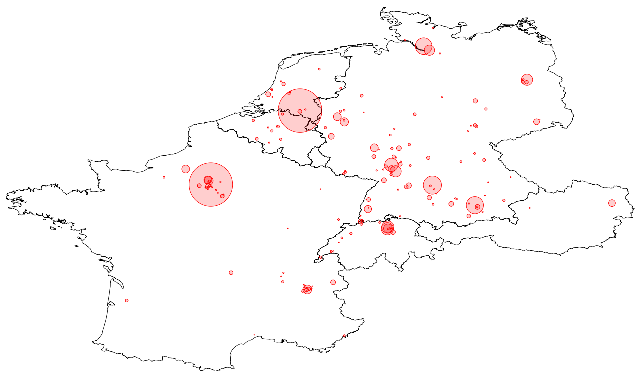
(a) 1990-1993



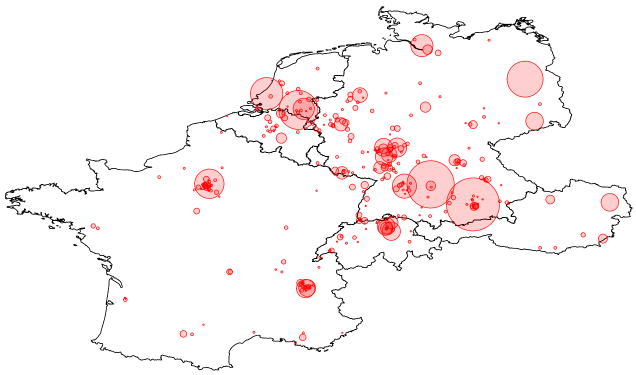
(b) 1993-1996



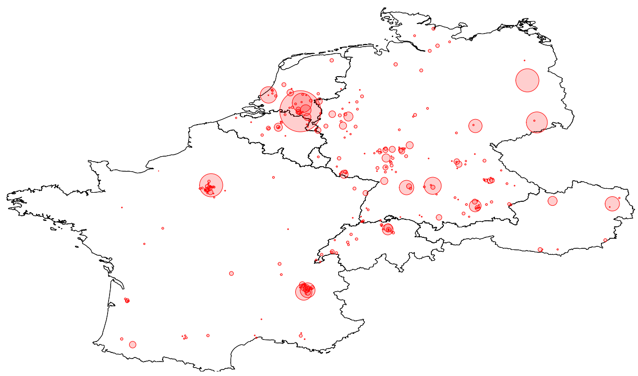
(c) 1996-1999



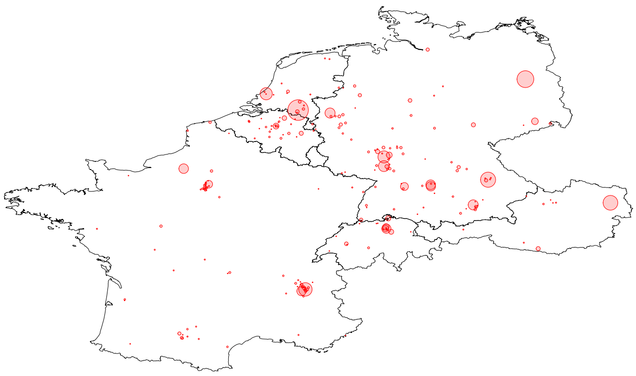
(d) 1999-2002



(e) 2002-2005



(f) 2005-2008



(g) 2008-2011

Figure A.3: Nanotechnology patent activity in Europe over time.

APPENDIX B

Kd-function elaboration

In this appendix the usage of the Kd-function in this thesis is shortly discussed and an example is given of how the Kd-function works in practice.

Repeating equation 4.1 from part I in this thesis,

$$\widehat{K}(d) = \frac{1}{n(n-1)h} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} f\left(\frac{d-d_{ij}}{h}\right) \tag{4.1}$$

The Kd-function calculates the Euclidian distance between n points, represented by d_{ij}. In this equation a Gaussian function (representing the normal distribution) is used for f and Silverman's 'rule of thumb' is used for the bandwidth h. Distances d are within a specific chosen interval $[0,u]$.

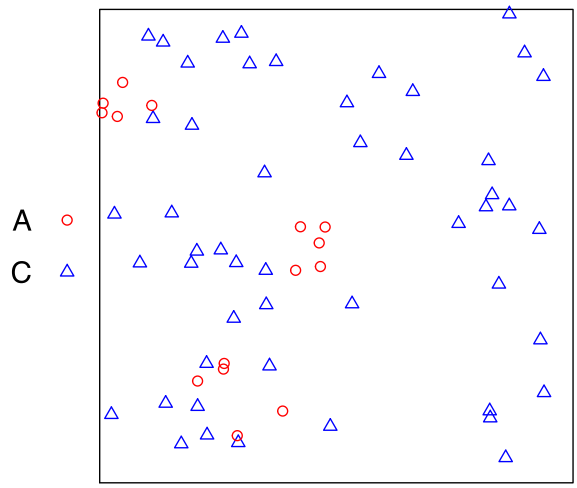


Figure B.1: 10 by 10 window with 15 clusterd points (A) and 50 random distributed points (C)

In order to see how this function behaves, in figure B.1 clustered points (A) are plotted in a 10 by 10 window. These points are deliberately divided over three clusters with a normal distribution ($\sigma = 0.5$). To simulate overall concentration of innovation, 50 points (C) are randomly distributed in the window by a uniform distribution.

The random distributed points are used to control for the overall population as wel as to create counterfactuals for the calculation of the confidence intervals. In each simulation of the monte carlo analysis, 15 points are randomly drawn from the control points after which the set of bilateral distances is calculated. If we run 1000 simulations and set the interval at [0,10] (the window size), the Kd-function of the clusterd points with its global confidence intervals results in the graph seen in figure B.2.

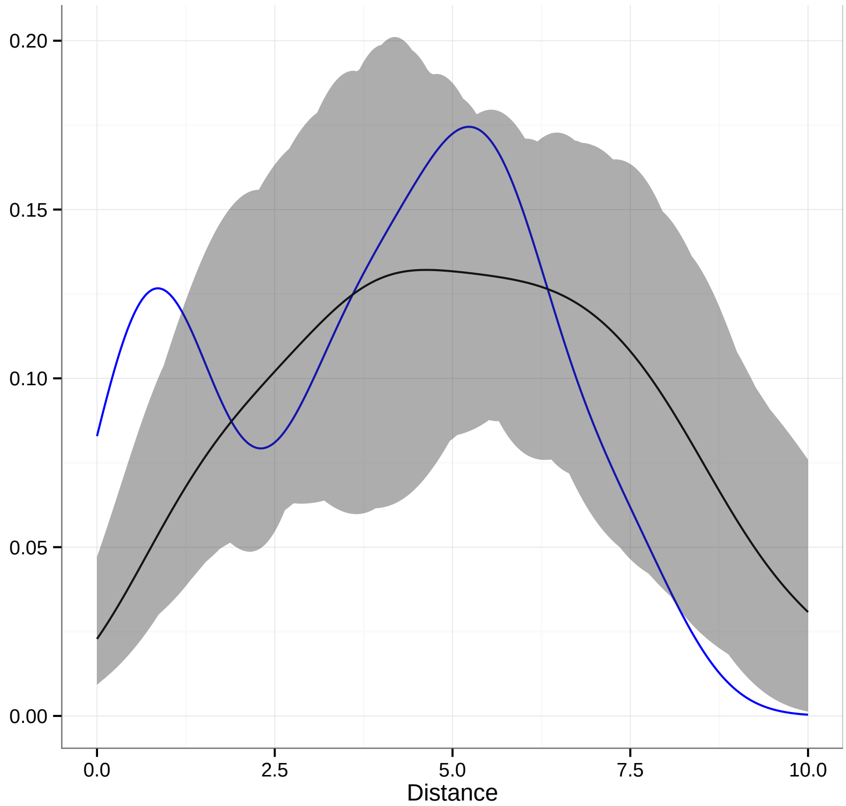


Figure B.2: 10 by 10 window with 15 clusterd points (A) and 50 random distributed points (C)

If we analyze this figure, we first see that group A is significantly localized at a short distance (0-1.5). This makes sense of course, as group A is deliberatly clustered with a normal distribution with a standard deviation of 0.5. At the larger distances, we see a significant dispersion. This is phenonomenon is inherently connected to the Kd-function, as localization at short distances results in dispersion at greater distances (the function always sums to 1). Therefore, normally we would choose a smaller interval for the kd-function, as the distance of 10 is not very relevant to analyze here. We can also see that the confindence band is quite large in width, wich can be explained by the relatively low amount of spatial points used in this example.

The clustering index used in this thesis can be calculated by summing the area between the Kd-function of the analyzed group of points and the upper confidence band:

$$CI_T = \sum_{d=0}^{d=u} \widehat{K_T}(d) - \overline{\overline{K_T}}(d) \tag{4.2}$$

In this example the comparable clustering index is:

$$CI_T = 17.768$$

However, the height of CI depends on the precision scale that is used for the distances in the interval $[0,u]$. In a comparison, the relative size of the clustering index is the same as this scale is kept at the same precision for all calculations.

Bibliography

Acs, Z., Anselin, L., & Varga, A. (2002). Patents and innovation counts as measures of regional production of new knowledge. *Research policy*, *31*, 1069–1085.

Acs, Z. J., Audretsch, D. B., & Feldman, M. P. (1992). Real Effects of Academic Research : Comment. *The American Economic Review*, *82*(1), 363–367.

Agarwal, R. & Audretsch, D. B. (2001). Does Entry Size Matter? The Impact of the Life Cycle and Technology on Firm Survival *. *The Journal of Industrial Economics*, *XLIX*(1), 21–43.

Andersen, B. (1999). The hunt for S -shaped growth paths in technological innovation : a patent study. *Journal of Evolutionary Economics*, *9*, 487526.

Arbia, G., Espa, G., & Quah, D. (2007). A class of spatial econometric methods in the empirical analysis of clusters of firms in the space. *Empirical Economics*, *34*(1), 81–103.

Archibugi, D. & Pianta, M. (1996). Measuring technological change through patents and innovation surveys. *Technovation*, *16*(9), 451–468.

Asheim, B. & Gertler, M. (2005). The geography of innovation. *The Oxford handbook of innovation*, 291–317.

Asheim, B. r. T., Boschma, R., & Cooke, P. (2011). Constructing Regional Advantage: Platform Policies Based on Related Variety and Differentiated Knowledge Bases. *Regional Studies*, *45*(7), 893–904.

Asheim, B. r. T. & Coenen, L. (2005). Knowledge bases and regional innovation systems: Comparing Nordic clusters. *Research Policy*, *34*(8), 1173–1190.

Audretsch, D. B. & Feldman, M. P. (1996a). R & D Spillovers and the Geography of Innovation and Production. *The American Economic Review*, *86*(3), 630–640.

Audretsch, D. B. & Feldman, M. P. (2004). Knowledge Spillovers and the Geography of Innovation. In J. V. Henderson & J.-F. Thisse (Eds.), *Handbook of Regional and Urban Economics*, volume 4 (pp. 2713–2739).

Audretsch, D. B. & Feldman, P. (1996b). Innovative Clusters and the Industry Life Cycle *. *Review of Industrial Organization*, *11*, 253–273.

Bathelt, H. & Peng-Fei, L. (2014). Evolutionary Economic Geography and Relational Geography. In M. M. Fischer & P. Nijkamp (Eds.), *Handbook of Regional Science* (pp. 591–607). Springer Berlin Heidelberg.

Beaudry, C. & Schiffauerova, A. (2009). Who's right, Marshall or Jacobs? The localization versus urbanization debate. *Research Policy*, *38*(2), 318–337.

Boschma, R. (2005). Proximity and Innovation: A Critical Assessment. *Regional Studies*, *39*(1), 61–74.

Boschma, R., Balland, P.-a., & Kogler, D. (2011). A relational approach to knowledge spillovers in biotech. Network structures as drivers of inter-organizational citation patterns. *Papers in Evolutionary Economic Geography*.

Boschma, R., Balland, P.-a., & Kogler, D. F. (2014). Relatedness and technological change in cities: the rise and fall of technological knowledge in US metropolitan areas from 1981 to 2010. *Industrial and Corporate Change*, 1–28.

Boschma, R. & Frenken, K. (2006). Why is economic geography not an evolutionary science? Towards an evolutionary economic geography. *Journal of Economic Geography*, *6*(3), 273–302.

Boschma, R. & Frenken, K. (2011a). Technological relatedness, related variety and economic geography. In P. Cooke, B. T. Asheim, & R. Boschma (Eds.), *Handbook of Regional Innovation and Growth* (pp. 187–197). Edward Elgar Publishing.

Boschma, R. & Frenken, K. (2011b). The emerging empirics of evolutionary economic geography. *Journal of Economic Geography*, *11*(2), 295–307.

Boschma, R. & Iammarino, S. (2009). Related variety, trade linkages, and regional growth in Italy. *Economic Geography*, *85*(3), 289–311.

Boschma, R. & Martin, R. (2010). The aims and scope of evolutionary economic geography. In R. Boschma & R. Martin (Eds.), *Handbook of Evolutionary Economic Geography* (pp. 3–39).

Boschma, R. & Wenting, R. (2007). The spatial evolution of the British automobile industry: Does location matter? *Industrial and Corporate Change*, *16*(2), 213–238.

Braunerhjelm, P. & Borgman, B. (2006). Agglomeration, Diversity and Regional Growth: The effects of poly-industrial versus mono-industrial agglomerations.

Breschi, S. (1999). Spatial patterns of innovation: evidence from patent data. In *The organization of economic innovation in Europe* (pp. 71–102).

Breschi, S. (2000). The Geography of Innovation : A Cross-sector Analysis. *Regional Studies*, *34*(3), 213–229.

Breschi, S. (2011). The geography of knowledge flows. In P. Cooke & B. r. T. Asheim (Eds.), *Handbook of Regional Innovation and Growth* (pp. 132–142). Edward Elgar Publishing.

Buenstorf, G. & Klepper, S. (2009). Heritage and Agglomeration: The Akron Tyre Cluster Revisited. *The Economic Journal*, *119*(537), 705–733.

Carrincazeaux, C. & Coris, M. (2011). Proximity and innovation. In P. Cooke & B. r. T. Asheim (Eds.), *Handbook of Regional Innovation and Growth* (pp. 269–281). Edward Elgar Publishing.

Cohen, W. M. & Levinthal, D. A. (1990). Absorptive Capacity : A New Perspective on and Innovation Learning. *Administrative Science Quarterly*, *35*(1), 128–152.

Cowan, R., Jonard, N., & Özman, M. (2004). Knowledge dynamics in a network industry. *Technological Forecasting and Social Change*, *71*(5), 469–484.

CPC (2014). http://www.cooperativepatentclassification.org/about.html, Date Accessed: 15-10-2014.

Crescenzi, R. (2014). Changes in Economic Geography Theory and the Dynamics of Technological Change. In M. M. Fischer & P. Nijkamp (Eds.), *Handbook of Regional Science* (pp. 649–666). Springer Berlin Heidelberg.

Deyle, H.-G. & Grupp, H. (2005). Commuters and the regional assignment of innovative activities: A methodological patent study of German districts. *Research Policy*, *34*(2), 221–234.

Dosi, G. & Nelson, R. R. (1994). Evolutionary Economics An introduction to evolutionary theories in economics *. *Journal of Evolutionary Economics*, *4*, 153–172.

Dumais, G., Ellison, G., & Glaeser, E. L. (2002). Geographic Concentration as a Dynamic Process. *The Review of Economics and Statistics*, *84*(2), 193–204.

Duranton, G. & Overman, H. G. (2005). Testing for Localization Using Micro-Geographic Data. *Review of Economic Studies*, *72*(4), 1077–1106.

Duranton, G. & Puga, D. (2001). Nursery Cities : Urban Diversity , Process Innovation , and the Life Cycle of Products. *The American Economic Review*, *91*(5), 1454–1477.

Ellison, B. G., Glaeser, E. L., & Kerr, W. R. (2010). What Causes Industry Agglomeration ? Evidence from Coagglomeration Patterns. *American Economic Review*, *100*(June), 1195–1213.

Ellison, G., Glaeser, E. L., Journal, S., & October, N. (1997). Geographic Concentration in U . S . Manufacturing Industries : A Dartboard Approach. *Journal of Political Economy*, *105*(5), 889–927.

Florida, R. (2005). The world is spiky. Globalization has changed the economic playing field, but hasn't levelled it.

Frenken, K., Van Oort, F., & Verburg, T. (2007). Related Variety, Unrelated Variety and Regional Economic Growth. *Regional Studies, 41*(5), 685–697.

Gaigné, C. & Thisse, J. (2014). New Economic Geography: Past and Future. In M. M. Fischer & P. Nijkamp (Eds.), *Handbook of Regional Science* (pp. 539–568). Springer Berlin Heidelberg.

Griliches, Z. (1998). *Patent Statistics as Economic Indicators : A Survey*, volume I.

Groot, H. L. F. D., Poot, J., & Smit, M. J. (2007). Agglomeration , Innovation and Regional Development. *Tinbergen Institute Discussion Paper*.

Hall, B. H. (2005). Exploring the Patent Explosion. *Journal of Technology Transfer*, *30*(1/2), 35–48.

Hartog, M., Boschma, R., & Sotarauta, M. (2012). The Impact of Related Variety on Regional Employment Growth in Finland 1993-2006: High-Tech versus Medium/Low-Tech. *Industry & Innovation*, *19*(6), 459–476.

Heimeriks, G. & Boschma, R. (2013). The path- and place-dependent nature of scientific knowledge production in biotech 1986-2008. *Journal of Economic Geography*, 1–26.

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

Horvath, M., Schivardi, F., & Woywode, M. (2001). On industry life-cycles: delay, entry, and shakeout in beer brewing. *International Journal of Industrial Organization*, *19*(7), 1023–1052.

Howells, J. R. L. (2002). Tacit Knowledge , Innovation and Economic Geography. *Urban Studies*, *39*(5-6), 871–884.

Jaffe, A. B. (1989). Real Effects of Academic Research. *The American Economic Review*, *79*(5), 957–970.

Karlsson, C. & Grå sjö, U. (2014). Knowledge Flows, Knowledge Externalities, and Regional Economic Development. In M. M. Fischer & P. Nijkamp (Eds.), *Handbook of Regional Science* (pp. 413–437). Springer Berlin Heidelberg.

Klepper, S. (1996). Entry , Exit , Growth , and Innovation over the Product Life Cycle. *The American Economic Review*, *86*(3), 562–583.

Klepper, S. (1997). Industry Life Cycles. *Industrial and Corporate Change*, *6*(1), 145–182.

Klepper, S. (2007). Disagreements, Spinoffs, and the Evolution of Detroit as the Capital of the U.S. Automobile Industry. *Management Science, 53*(4), 616–631.

Krugman, P. (1991a). *Geography and Trade*. Leuven University Press-MIT Press, Cambridge, MA.

Krugman, P. (1991b). Increasing Returns and Economic Geography. *The Journal of Political Economy*, *99*(3), 483–499.

Leydesdorff, L., Alkemade, F., Heimeriks, G., & Hoekstra, R. (2014). Geographic and Technological Perspectives on "Photovoltaic Cells :" Patents as Instruments for Exploring Innovation Dynamics.

Leydesdorff, L. & Bornmann, L. (2012). Mapping (USPTO) patent data using overlays to Google Maps. *Journal of the American Society . . .* , *63*(7), 1442–1458.

Leydesdorff, L. & Rafols, I. (2011). Indicators of the interdisciplinarity of journals: Diversity, centrality, and citations. *Journal of Informetrics*, *5*(1), 87–100.

Lu, K. (2009). A Study of Engineering Freshmen Regarding Nanotechnology Understanding. *Journal of STEM Education*, *10*(1&2), 7–16.

Maggioni, M. A. (1999). *Clustering Dynamics and the Location of High-Tech Firms*. PhD thesis.

Maine, E. & Garnsey, E. (2006). Commercializing generic technology: The case of advanced materials ventures. *Research Policy*, *35*(3), 375–393.

Malecki, E. J. (2014). The Geography of Innovation. In M. M. Fischer & P. Nijkamp (Eds.), *Handbook of Regional Science* (pp. 375–389). Berlin, Heidelberg: Springer Berlin Heidelberg.

Malmberg, A. & Maskell, P. (2002). The elusive concept of localization economies: towards a knowledge-based theory of spatial clustering. *Environment and Planning A*, *34*(3), 429–449.

Malmberg, A. & Maskell, P. (2010). An evolutionary approach to localized learning and spatial clustering. In R. Boschma & R. Martin (Eds.), *Handbook of Evolutionary Economic Geography* (pp. 391–405).

Malmberg, A., Sölvell, O., & Zander, I. (1996). Spatial Clustering, Local Accumulation of Knowledge and Firm Competitiveness. *Geograifska Annaler. Series B, Human Geography*, *78*(2), 85–97.

Marcon, E. & Puech, F. (2010). Measures of the geographic concentration of industries: improving distance-based methods. *Journal of Economic Geography*, *10*(5), 745–762.

Marcon, E. & Puech, F. (2011). A Typology of Distance-Based Measures of Spatial Concentration, 1–14.

Marshall, A. (1890). *Principles of Economics*. MacMillan and Co., London, UK.

Martin, R. & Sunley, P. (1996). Paul Krugman ' s Geographical Economics and Its Implications for Regional Development Theory: A Critical Assessment. *Economic Geography*, *72*(3), 259–292.

Maurel, F. & Sédillot, B. (1999). A measure of the geographic concentration in french manufacturing industries. *Regional Science and Urban Economics*, *29*(5), 575–604.

McCann, B. T. & Folta, T. B. (2011). Performance differentials within geographic clusters. *Journal of Business Venturing*, *26*(1), 104–123.

Mccann, P. (2014). Schools of Thought on Economic Geography, Institutions, and Development. In M. M. Fischer & P. Nijkamp (Eds.), *Handbook of Regional Science* (pp. 527–538). Berlin, Heidelberg: Springer Berlin Heidelberg.

Mehta, M. D. (2002). Nanoscience and Nanotechnology: Assessing the Nature of Innovation in These Fields. *Bulletin of Science, Technology & Society*, *22*(4), 269–273.

Neffke, F., Henning, M., Boschma, R., Lundquist, K.-J., & Olander, L.-O. (2011). The Dynamics of Agglomeration Externalities along the Life Cycle of Industries. *Regional Studies*, *45*(1), 49–65.

Nelson, R. R. & Winter, S. G. (1982). *An evolutionary theory of economic change*. Harvard University Press.

Nooteboom, B. (2000). Learning by Interaction : Absorptive Capacity , Cognitive Distance and Governance. *Journal of Management and Governance*, *4*, 69–92.

OECD (2009). OECD Patent Statistics Manual. Technical report, OECD.

Ó Huallacháin, B. & Lee, D.-S. (2011). Technological Specialization and Variety in Urban Invention. *Regional Studies*, *45*(1), 67–88.

Oinas, P. & Malecki, E. J. (2002). The Evolution of Technologies in Time and Space: From National and Regional to Spatial Innovation Systems. *International Regional Science Review*, *25*(1), 102–131.

Palan, N. (2010). Measurement of Specialization - The Choice of Indices.

Press, K. (2006). *A Life Cycle for Clusters?: The Dynamics of Agglomeration, Change, and Adaption*. Springer.

Rao, C. R. (1982). Diversity and dissimilarity coefficients: a unified approach. *Theoretical Population Biology*, *21*, 24–43.

Rigby, D. L. (2013). Technological Relatedness and Knowledge Space: Entry and Exit of US Cities from Patent Classes. *Regional Studies*, (October 2014), 1–16.

Ripley, B. D. (1976). The Second-Order Analysis of Stationary Point Processes. *Journal of Applied Probability*, *13*(2), 255–266.

Romer, P. M. (1986). Increasing Returns and Long-Run Growth. *Journal of Political Economy*, *94*(5), 1002–1037.

Rothaermel, F. T. & Thursby, M. (2007). The nanotech versus the biotech revolution: Sources of productivity in incumbent firm research. *Research Policy*, *36*(6), 832–849.

Scheu, M., Veefkind, V., Verbandt, Y., Galan, E. M., Absalom, R., & Förster, W. (2006). Mapping nanotechnology patents: The EPO approach. *World Patent Information*, *28*(3), 204–211.

Scott, A. J. (2000). Economic geography: the great half-century. *Cambridge Journal of Economics*, *24*(4), 483–504.

Shapira, P., Youtie, J., & Kay, L. (2011). National innovation systems and the globalization of nanotechnology innovation. *The Journal of Technology Transfer*, *36*(6), 587–604.

Simmie, J. & Martin, R. (2010). The economic resilience of regions: towards an evolutionary approach. *Cambridge Journal of Regions, Economy and Society*, *3*(1), 27–43.

Sorenson, O., Rivkin, J. W., & Fleming, L. (2006). Complexity, networks and knowledge flow. *Research Policy*, *35*, 994–1017.

Steinle, C. & Schiele, H. (2002). When do industries cluster ? A proposal on how to assess an industry's propensity to concentrate at a single region or nation. *Research Policy*, *31*, 849–858.

Stirling, A. (2007). A general framework for analysing diversity in science, technology and society. *Journal of the Royal Society, Interface / the Royal Society*, *4*(15), 707–19.

Taniguchi, N. & Others (1974). On the basic concept of nanotechnology. In *Proc. Intl. Conf. Prod. Eng. Tokyo, Part II, Japan Society of Precision Engineering*, (pp. 18–23).

Utterback, J. M. & Suárez, F. F. (1993). Innovation, competition, and industry structure. *Research Policy*, *22*(1), 1–21.

Von Thünen, J. (1826). Der isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie. Hamburg: F. Perthes.

Weber, A. (1909). *Über den Standort der Industrien*. Tubingen: J.C.B. Mohr.