THE GEOGRAPHICAL AND TECHNOLOGICAL ORIGINS OF RADICAL INNOVATION

Copernicus Institute of Sustainable Development

Department Innovation, Environment and Energy Sciences

Faculty of Geosciences| Utrecht University

Guus Hutschemaekers
g.m.hutschemaekers@students.uu.nl
Student ID: 4048318

Supervisor
Gaston Heimeriks – Assistant Professor
g.j.heimeriks@uu.nl

Second Reader:
Foortje Alkemade - Professor
f.alkemade@tue.nl

7 October 2015

SUMMARY

The purpose of this study was to gain insight in the technological and geographical origins of radical innovation. Theory states that radical innovations are recombinations of technologically unfamiliar knowledge, which are more likely to be conceived in an environment of high geographical proximity. In this research, I investigate the relationship between regional knowledge diversity and radical innovation potential, and the mechanisms behind this relation. I calculated a radicality measure for nearly 50 million patents based on the coherence of their technology classes, and compared this to the regional knowledge diversity of 1400 TL2 and TL3 OECD regions based on 5 million region-assigned patents. The results show that there is indeed a positive relationship between regional knowledge diversity and radical innovation potential. Furthermore the positive influence of technological and geographical proximity on knowledge spillovers is confirmed, as is dependency of an innovation on its evolutionary knowledge base. Finally, the results indicate that, in opposition to theory, radical innovations do not have an increased potential of success, whereas the increased risk of failure was confirmed. This implies that the theoretical beliefs regarding the evolutionary origins of radical innovation can be confirmed. The presumed societal impact of these radical innovations on the contrary, should be questioned, as no proof of increased impact is found in the present study.

ACKNOWLEDGEMENTS

Firstly I would like to thank dr. Gaston Heimeriks for his supervision, input and guidance throughout the period of my master's thesis. In addition, my thanks goes out to prof. Floortje Alkemade, for the critical feedback on my research proposal and for taking the time to read and assess my thesis. I also want to thank dr. Ad Feelders for reviewing my data and methodology. Finally I would like to thank my fellow students and friends, in particular Matthee le Cointre, Pieter van Os, Anne van Wijngaarden, Rudolf Burggraaf and Jessy Hendriks, for their enlightening insights, sincere feedback and for supporting me throughout the process of this thesis.

TABLE OF CONTENTS

1 INTRODUCTION

1.1 BACKGROUND

Radical innovation is said to be the driving force of technological, industrial and societal change, and its importance to society is widely recognized (Arts & Veugelers, 2013; Schoenmakers & Duysters, 2010).

Innovation in general is described by Schumpeter (1942) as a new combination of existing knowledge, or 'neue Kombinationen'. The topic of innovation has received great attention from scholars of many disciplines, regarding its definition, origins, implications and characteristics. One of these topics is the nature of innovation, where innovation is classified as *radical* versus *incremental*. These classifications of innovation are defined in numerous fashions throughout the literature. Some classifications emphasize the role of technology, where others focus on the market implications (Garcia & Calantone, 2002).

In this study, radical innovations are approached from a technological perspective. In this context, radical innovations have high novelty value and build on uncommon combinations of knowledge, whereas incremental innovations, make improvements to existing artefacts. Radical innovations are considered to account for a large share of *disruptive* innovations, i.e. innovations with high impact on society, disrupting market, technology or industry standards (Freeman, 1992; Tidd, Bessant, & Pavitt, 2005; Winter & Nelson, 1982). At the same time, radical innovations come with an elevated risk of failure (Andersen, 2014; Keizer & Halman, 2009; Uzzi, Mukherjee, Stringer, & Jones, 2013). Because of their potential *impact* on society, it is essential to understand the origins and mechanisms of radical innovation.

Many studies have contributed to the understanding of the origins of radical innovation, particularly from firm perspective (Ahuja & Lampert, 2001; Chandy & Tellis, 2000) and inventor perspective (Singh & Fleming, 2010). An ongoing discussion regarding the sources of radical innovation has been on the role of young entrepreneurial firms versus that of the established incumbents (Arts & Veugelers, 2013; Schoenmakers & Duysters, 2010). While some argue that incumbent firms lack the ability to innovate radically (the *incumbents* curse), others dispute the validity of this claim, and stress the important role incumbents play in radical and disruptive innovation (Chandy & Tellis, 2000).

Arts and Veugelers (2013), argue that the 'confusion' around this *incumbents curse* is caused by a lack of understanding of the origins of radical innovations, as little empirical research is available on the technological origins of radical innovations. Similar conclusions were drawn by Schoenmakers and Duysters (2010), who argue that "large-scale empirical studies into the technological origins of radical innovations are sparse and almost non-existing" (p.1051). Arts and Veugelers (2013) and Schoenmakers and Duysters (2010) made an effort to address this gap. They found that radical innovations strongly rely on highly diverse existing knowledge.

In order to understand the mechanisms behind radical innovation, one first needs to understand the mechanisms of the innovation process in general. Many scholars have studied this issue over the past decades, and a rich body of literature is available on this matter. One of the most important mechanisms of the innovation process is knowledge spillovers, the exchange of ideas between firms, institutions or individuals (Aldieri, 2011; Lychagin, Pinkse, Slade, & Reenen, 2010; Marrocu, Paci, & Usai, 2014; Maurseth & Verspagen, 2002). In this evolutionary process, compatible bits of knowledge, from different sources are combined to create new knowledge or technologies (Heimeriks & Boschma, 2013; Iammarino & McCann, 2006).

Conditions and forces enabling or fostering the diffusion of ideas through knowledge spillovers are regarded important to the innovation process (Aldieri, 2011). An essential condition for knowledge spillovers is that of *technological* or *cognitive proximity*. High technological proximity is found to have a strong positive influence in knowledge spillovers (Aldieri, 2013; Jaffe, Trajtenberg, & Henderson, 1993; Marrocu et al., 2014; Verspagen, 2006).

Yet the high novelty value of radical innovation is not achieved by combinations of similar knowledge, but on the contrary, is supported by diverse knowledge sources, with high technological distance between the building blocks of innovation (Nooteboom, Van Haverbeke, Duysters, Gilsing, & van den Oord, 2007). Hence, radical innovation cannot benefit from the advantageous conditions of high technological proximity. Fortunately, technological proximity is not the only favourable condition for innovation. Boschma (2005) identifies a series of proximities, of which geographical proximity is believed to positively affect knowledge spillovers. Marrocu et al. (2014) even argue geographical proximity can partially compensate for lack of technological proximity.

Based on these theories, it follows that environments with high geographical proximity and high knowledge diversity should be fertile environments for the radical innovation process. A tangible approach to express high geographical proximity, as occasionally used in research, is the approach of regions (Boschma, 2005; Ponds, van Oort, & Frenken, 2007). An important advantage of this approach is the level of standardisation, enhancing the amount of available data and comparability with other studies. So a region with a diverse portfolio of knowledge, is an environment with high geographical proximity providing knowledge sources with high inter-knowledge technological distance. Hence seems straightforward that regions with a diverse knowledge base foster radical innovation, by enabling spillovers of diverse knowledge.

Yet (large scale) empirical support is lacking thus far. Furthermore, apart from the works by Arts and Veugelers (2013) and Schoenmakers and Duysters (2010), evolutionary technological attributes of radical innovation have received little attention in academic literature. Likewise, the relation between regional knowledge diversity and radical innovation has been barely touched upon in literature. Enhanced insights in the relation between technological diversity of a region, and its capability to innovate radically, are desirable as they can provide guidance in (regional) policy making regarding diversification, specialisation and expansion strategies. Such strategies may have important implications for a region's radical innovation potential, and therewith its potential to a podium for the frontrunners in technological development.

1.2 Research question

From the specified gap in prevalent literature, the following question follows:

Does regional knowledge diversity stimulate radical innovation?

Concepts constituting the research question will be introduced in section 2, by a thorough review of relevant literature captured in a theoretical framework. The highlights of this framework are expressed in hypotheses. Section 3 elaborates on how a comprehensive patent analysis is applied to test the hypotheses, after which the results confirm the positive effect of regional knowledge diversity on radical innovation in section 4. Section 5 discusses the implications and limitations and the final conclusions are presented in section 6.

2 THEORETICAL FRAMEWORK

2.1 RADICAL INNOVATION

Because of its widely recognised relevance, innovation is approached from many angles in the large, multidisciplinary body of literature on innovation. The definition of innovation is a topic of discussion to which papers have been dedicated (i.e. Baregheh, Rowley, & Sambrook, 2009).

The perspective adopted in this study, relies on the claim that the ultimate source of novelty and (radical) innovation is *recombination of existing knowledge* (Fleming, 2001; Neffke, 2009), or 'neue Kombinationen' (Schumpeter, 1942). This approach of innovation as a recombination of existing knowledge is widely accepted (Fleming, 2001; Utterback, 1996). Hargadon (2003), for example, refers to innovation as "*the recombination of elements of existing technologies"* (p. 68). Similarly Winter and Nelson (1982, p. 130) state in their economic approach, that innovations rely for a substantial amount on recombination of existing materials and concepts. In this evolutionary approach to knowledge production, it is argued that the that the building blocks of knowledge are often closely related, reflecting the path dependent nature of innovation (Arts, 2012; Heimeriks & Boschma, 2013).

These new recombinations of familiar knowledge or technologies are expected to result in a majority of small improvements of technology commonly referred to as *incremental innovation.* On the counterpole, unique combinations of previously unrelated knowledge, or *radical innovation* occurs only sporadically. From this remark, the first hypothesis follows:

H1: The majority of innovations is incremental

Even though the term radical innovation is covered in a number of definitions, Garcia & Calantone (2002) find in a comprehensive literature review that radical innovations are consistently modelled as discontinuities on technological or marketing factors. The perspective adopted in this study is that radicality is determined by the technological characteristics of an innovation: "*A radical product innovation is a new product that incorporates a substantially different core technology"* (Chandy & Tellis, 2000, p. 2) In this approach, innovations are classified as radical based on the high novelty value resulting from the combination of core technologies, as a result of a discontinuity on technological level. This demarcation of the concept *innovation* closely approaches the territory of *invention,* because the technological aspect of innovation is emphasized. Nevertheless this study aims to construct a framework elaborating on the technological innovation process, rather than the invention alone.

2.2 KNOWLEDGE DIVERSITY

Given that innovations are recombinations of knowledge, it can be argued that innovation radicality is reflected in the *diversity* of knowledge incorporated in the innovation. That is, combinations of closely related knowledge typically represent *incremental* innovation, whereas combinations of dissimilar knowledge represent radical innovation. From evolutionary perspective it follows naturally that the diversity of building blocks an innovation stems from is reflected in the diversity of knowledge incorporated in an innovation: innovation radicality. The second hypothesis is therefore:

> *H2: Radical innovations stem from more diverse knowledge sources than incremental innovations*

Where this proposition refers to the technological aspect of radical innovation, other approaches to the definition of *radicality* focus more on the impact an innovation has on the market and subsequent innovations (Freeman, 1992; Rosenberg, 1994), in line with the concept of *creative destruction*. This phenomena is also known as *breakthrough* or *disruptive innovation* (Ahuja & Lampert, 2001; C. M. Christensen & Raynor, 2003; Dahlin & Behrens, 2005; Fleming, 2001) and will be denoted as such in this study. Established technologies, organisations and even industries can be overturned by disruptive innovations (discontinuity), as a disruptive innovation can be a starting point of new products, technologies or services that make the existing ones inferior (Abernathy & Utterback, 1978; C. Christensen & Rosenbloom, 1995; Freeman, 1992; Utterback, 1996).

Due to their high novelty value, radical innovations have increased potential to be disruptive (Tidd et al., 2005). Yet only a small share of radical innovations gain sufficient momentum to take off. This is because novel combinations of knowledge are mostly unusual ideas. This means, that even if the idea has great potential, it may put off potential adopters, investors or other actors in the "environment" because it is unfamiliar. This high-risk/high-gain property of radical innovations is widely recognised (Andersen, 2014; Keijl, 2011; Keizer & Halman, 2009; Schilling & Green, 2011; Uzzi et al., 2013), and forms the basis for the third hypothesis:

> *H3: Radical innovations have high-risk/high-gain characteristics in terms of impact*

2.3 Knowledge spillovers

Having established that radical innovation is constituted by new combinations of existing but previously unconnected knowledge, the next question is how these eccentric combinations occur. A number of studies on radical innovation show cross-industry knowledge combinations are likely to constitute radical innovations (Neffke, 2009). Such spillovers of knowledge between different industries or organisations are important mechanisms through which radically new combinations occur.

The concept of knowledge spillovers in this study is defined as the exchange of ideas and knowledge between individuals, the an essential source of innovation (Carlino, 2001). Likewise, Jaffe (1986) identifies (R&D) spillovers as an important factor from the "supply side" of innovation. While knowledge spillovers are considered risks of inter-firm collaborations, and can be rather undesirable from firm perspective, "*From a purely technological point of view, R&D spillovers constitute an unambiguous positive externality*" (Jaffe, 1986, p. 984). Spillovers occur accidental and intentional, in formal and informal settings. Whether or not spillovers happen intentionally, is beyond the scope of this research, as only the result, the inventive output, matters (Aldieri, 2011; Marrocu et al., 2014; Maurseth & Verspagen, 2002). Ultimately, knowledge spillovers are considered the primary mechanisms through which new combinations are conceived (Amin & Wilkinson, 1999; Boschma, 2005; Knoben & Oerlemans, 2006; Mackinnon, Cumbers, & Chapman, 2002).

2.4 Proximities

The conditions in which knowledge spillovers occur have been profoundly researched, and proximities are argued to have strong influence on these mechanisms of the innovation process. Boschma (2005) has outlined a set of five proximities that play a role in innovation. In line with (and with acknowledgment of) Boschma's review, Marrocu et al. (2014) found that for knowledge spillovers, technological and geographical proximity are by far most influential.

High technological proximity (or low technological distance) foster knowledge spillovers, because related knowledge sources have increased absorptive capacity, i.e. the ability to assimilate, use and apply knowledge (Boschma, 2005; Cohen & Levinthal, 1990; Knoben & Oerlemans, 2006). This is because mutual understanding increases with a more similar knowledge base of involved actors. Below a minimum level of technological proximity, knowledge spillovers are very unlikely to occur due to cognitive dissonance, and most spillovers will likely occur between related technologies (Cohen & Levinthal, 1990; Nooteboom et al., 2007).

The fourth and fifth hypothesis are therefore:

H4: The diversity of knowledge sources combined into innovations has a 'natural' maximum well below the theoretical maximum

H5: The majority of knowledge spillovers happen between related knowledge sources

Yet, as previously argued, racial innovations are not the result of combinations of familiar knowledge. In fact, radical innovations are the product of combinations of knowledge with high diversity or low inter-knowledge technological proximity. Hence, the benefit of technological proximity on spillovers is limited for radical innovations. Marrocu et al. (2014) found that low technological proximity can, at least partially, be compensated by high geographical proximity.

Geographical proximity is used to denote absolute distance between two collaborating actors (Boschma, 2005; Knoben & Oerlemans, 2006). Its most distinguishing property is its ability to facilitate face-to-face interactions, which fosters the transfer of rich (tacit) knowledge. It is seen as an important support for inter-organisational collaborations (Boschma, 2005; Knoben & Oerlemans, 2006; Oerlemans, Meeus, & Boekema, 2001). It is also strongly associated with both intended and unintended knowledge spillovers, resulting in 'autonomous' knowledge diffusion, therewith stimulating innovation (Jaffe, 1989; Marrocu et al., 2014).

A suitable approach to geographical proximity is the more tangible concept of *regions*. A region is in essence an delimited area of high geographical proximity. In innovation studies, regions are a popular unit of analysis because they can be unambiguously defined, and usually remain stable over time, enhancing the possibilities of inter-regional comparison. The argued influence of high geographical proximities on knowledge spillovers, is in line with the concept of *place dependency* (Heimeriks, Alkemade, Schoen, Villard, & Laurens, 2013; Heimeriks & Boschma, 2013). Since knowledge resources in regions are located close to each other, spillovers between these resources are promoted by high geographical proximity. Therefore, the sixth hypothesis predicts the reliance of knowledge production and innovation on locally available resources:

H6: Innovations rely for a significant part on regional knowledge

2.5 CONCEPTUAL MODEL

The key concepts of the elaborated theories were summarized in the conceptual model as shown in Figure 1. It shows how both *technological* and *geographical* proximity positively affect *knowledge spillovers*, the most important mechanism through which new combinations of knowledge occur. Yet technological proximity has a negative effect on the *diversity* of knowledge sources leading to an innovation. This *diversity is* quantified by the aggregated *technological distance* between the knowledge sources. The diversity of knowledge sources is reflected in the *radicality* of the resulting innovation, which in turn has a high-risk/high-gain relation with the impact of the innovation.

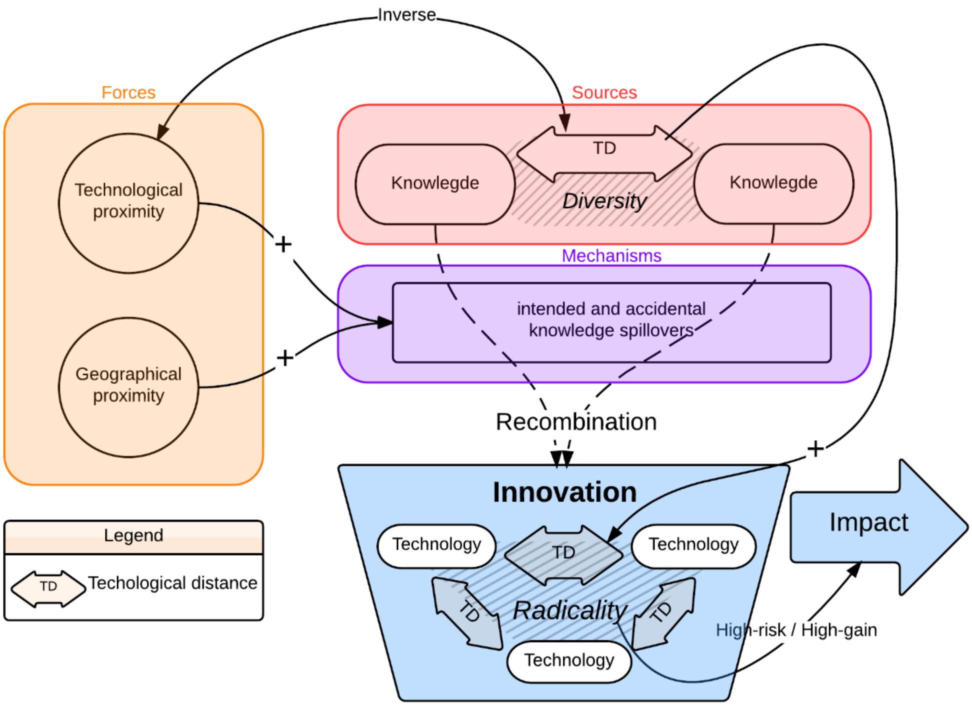


Figure 1 Conceptual model of theories

From this model follows a final overarching hypothesis:

> *H7: Highly radical innovations are more likely to emerge from regions with greater knowledge diversity.*

This hypothesis predominantly describes an expected relation between the overall diversity of knowledge residing in a region, and radicality of innovations originating from this region. This expected relation is based on the core-concepts of the model: Innovation as recombination of existing knowledge, knowledge spillovers as the primary mechanism behind innovation and the benefits of geographical and technological proximities. The high-risk/high-gain proposition highlights the societal impact.

3 METHODOLOGY

In this study, the core concepts are operationalized in patent data: Innovations are measured by patents, regional knowledge is defined by the patent portfolio of the region, and citations are assumed to be the result of knowledge spillovers. While this is clearly not a one on one representation of reality, patent data offer an extraordinary rich, structured source of information on the technological development, and are commonly used to measure (technological) innovation (Aldieri, 2013; Leydesdorff, Alkemade, Heimeriks, & Hoekstra, 2015) and knowledge spillovers (Maurseth & Verspagen, 2002).

3.1 DATA AND TOOLS

The patent (application) data was obtained from a number of databases. Over 80 million patent application records were retrieved from The Worldwide Patent Statistical Database (PATSTAT) by European Patent Office (EPO), version autumn 2014. These data included information on document numbers, application dates, priority status, citations, IPC-classifications and publication information.

This information was enriched with the Regional Patent database REGPAT[1] by OECD. The REGPAT data have been regionalised at a very detailed level, covering over 2000 regions in OECD countries, with cleaned harmonised applicants, inventors and corresponding addresses for over 5 million EPO/PCT patents. The data was connected to PATSTAT based on the application id's as assigned by EPO.

The next data source, the OECD Patent Quality Indicators database[2], contains a series of calculated patent indicators for EPO and USPTO patent applications (Squicciarini, Dernis, & Criscuolo, 2013). The forward citation counts in this database were used to measure impact. The information was joined with the PATSTAT and REGPAT information based on application id's.

In addition, the IFRIS technology classification was adapted. This is a refined technology classification of 35 technological fields and 389 technological subfields, derived from World Intellectual Property Organisation's (WIPO) International Patent Classification (IPC). These classifications are non-overlapping evenly distributed groups of technology classifications, optimised for calculating technological proximity (CIB, 2009; Schoen, Villard, & Larédo, 2014)

---

[1] OECD REGPAT database, February 2015

[2] OECD Patent Quality Indicators database, February 2015

The last patent data collected was the ECOOM-EUROSTAT-EPO PATSTAT Person Augmented Table (EEE-PPAT), created by ECOOM (K.U. Leuven) in collaboration with Sogeti, EPO and EUROSTAT (OECD), and was used for control variables. In this dataset, the type of patentee is recorded for the PATSTAT applications: Private business enterprise, university/higher education institution, governmental agencies and individuals.

All abovementioned data was uploaded and structured in Google BigQuery. This SQL like database allows for swift execution of queries over large amounts of data. Through the BigRQuery R-package (Wickham, 2015), the database was connected to R (R Core team, 2015) running with RStudio (RStudio Team, 2015) which was used to process and analyse the data.

Finally, regional statistics were obtained from the OECD statistics website, i.e. demographics and regional account data, to be used as control variables on region-level regressions.

3.2 Operationalisation

The rich collection of data needed narrowing and further processing to allow for testing the hypotheses. In the selection of measures and models to apply to the data, I relied strongly on publications of related studies.

In the paragraphs below, origins of all relevant variables will be discussed, to be finally summarized in Table 1: Operationalisation of variables.

Technological proximity

First, the 187 million records in PATSTATS application-IPC table were assigned an IFRIS technology class based on the concordance table[3] provided by IFRIS using BigQuery 01[4]. Next, the priority patents were filtered from the main application table, using the methodology as suggested by de Rassenfosse (2013). From these applications (about 39 million) a co-occurrence matrix of all 389 IFRIS subclasses was constructed from which the *technological relatedness* or proximity for every pair of was calculated using the Jaccard index, which quantifies the technological proximity as follows:

$$\varphi_{i,j} = \frac{occ_{ij}}{occ_i + occ_j - occ_{ij}} \qquad (1)$$

[3]The ifris-ipc class concordance table is in the datasources directory of the GitHub repository: Guus-H/thesis

[4] The queries used on the patent databases are available in the file 'BigQueries.sql' on the github repository

Where $\varphi_{i,j}$ is the technological proximity between technology i and j, occ_{ij} denotes the number of times technologies i and j occur together in a patent, and occ_i and occ_j denoting the occurrence of technologies i and j respectively. This method has proven its propriety in similar analyses by many others (Boschma, Heimeriks, & Balland, 2014; Gerken & Moehrle, 2012; G Thoma, Torrisi, Gambardella, & Guellec, 2010; Grid Thoma & Torrisi, 2007).

Patent radicality

The technological proximity values can be used to calculate a technological coherence of a group of classes, in this case all classes assigned to a single patent. This was achieved by taking the average of the proximity scores of every pair (φ_{ij}) of technologies for patent p (eq. 2), in line with the methodology as used by Kogler, Rigby and Tucker (2013). Patents with only one technology class were excluded from analysis as their technological coherence could not be calculated with the selected method.

$$Technological\ Coherence = \frac{\sum_{ij \in p}^{n} \varphi_{i,j}}{n} \tag{2}$$

Because the technological coherence is directly derived from technological proximity, it will return a value between 0 and 1 where values close to zero indicate rare combinations of classes in a patent and values closer to 1 indicate very common combinations. A technological innovation radicality index is operationalized as the inverse of the technological coherence: $Technological\ radicality = \frac{1}{Technological\ Coherence}$.

An important issue which has to be noted, is that radical innovations could potentially initiate new technological paradigms. Therewith, a novel combination could become common after its initial publication. If a single technology map is used to calculate the technological coherence of all patents, important game-changing innovations could be unjustly regarded non-radical. Therefore the technology proximity matrix used to calculate technological coherence and diversity was calculated for each year from 1980 up to 2012, including all data of priority patents applied for in the 10 years before the calculation year. This to distinguish “innovators” from “imitators”.

Knowledge diversity

The knowledge diversity was calculated for a single patent’s evolutionary knowledge base, and later for regions. The knowledge base of a patent, was operationalized as the set of backward citations of a single patent (Figure 2). The diversity of the knowledge base was calculated from the IFRIS technology classes occurring in the set of backwards citations, regardless of the patent they were assigned to. The same approach for the regional knowledge diversity was used: All IFRIS classes and their occurrence frequencies were counted for each region, to define a region technology portfolio (Figure 2), using the patents solely to connect the classes to the region.

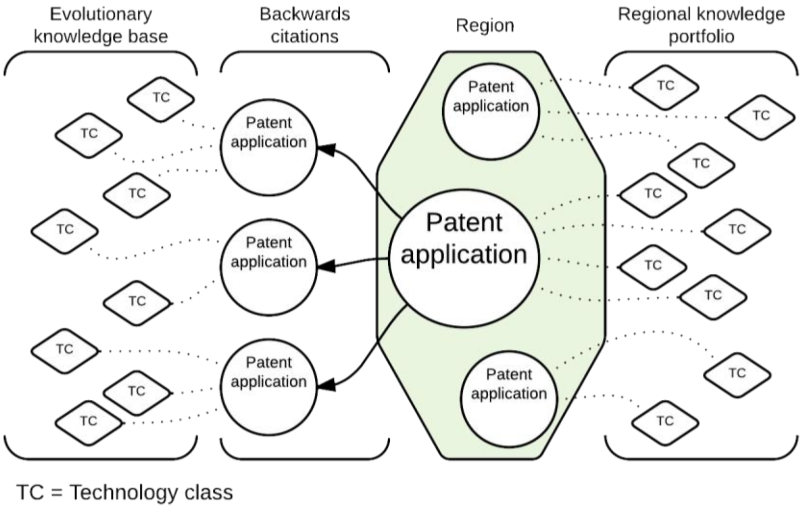


Figure 2 Schematic drawing of operationalisation knowledge base

The knowledge diversities, were calculated from the technological proximity values of the technology classes involved. Because the distribution of relatedness values was highly skewed, they were log transformed using a natural logarithm, after which the values were rescaled between 0 and 1. Finally, pairs of non-related technology classes were reassigned a relatedness value of 0, because these values were omitted in the log-transformation.

After this redistribution of relatedness values, the Rao-Stirling diversity was calculated, as proposed by Stirling (2007, p. 712). This measure has been used by Leydesdorff, Kushnir and Rafols (2014) amongst others, to estimate technology diversity based on patent classes. It is defined as the sum of pairwise disparities for the technology classes, weighed in proportion to the relative importance of a technology class in the set. (eq. 3)

$$Diversity\ \Delta = \sum_{ij(i \neq j)} \left(1 - \overline{\varphi_{\mathrm{i,j}}}\right) \cdot p_i \cdot p_j \qquad (3)$$

Where p_i and p_j are relative occurrences of technology i and j in a given set ($p_i = \frac{occ_i}{N_{occ}}$) and $\overline{\varphi_{\mathrm{i,j}}}$ is the corrected technological proximity for technology i and j. The use of this measure allowed assigning a measure of knowledge diversity to any set of patents in a single indicator, while accounting for variety, balance and disparity, and eliminating the need for further correction for set size (Stirling, 2007).

The knowledge diversity has been calculated for the backwards citations of individual patents (knowledge base) and for geographical regions. Calculating the knowledge diversity of backward citations is a resource-intensive operation. For each patent the technology classes of all backward citations were looked up, and from this the diversity was calculated. To limit the number of required lookups and computations a random sample of 100,000 REGPAT patents was selected, for which the knowledge diversity of backwards citations was calculated.

Regions and regionality

In this study, geographical proximity was operationalized through regions. The Territorial Level 2 and 3 (TL2 and TL3) regions, as defined by the OECD have been adopted. Patent-region information was obtained from the inventor-address tables of REGPAT. Inventor addresses were selected rather than applicant addresses as inventor addresses are more likely to be close to the place of the actual invention because applicant addresses often reflect the location of a firm's headquarters.

For each TL2 and TL3 region with at least 1000 patent applications between 1980 and 2012 recorded in the REGPAT database, the regional knowledge diversity was calculated.

In addition the regionality of patents occurring in the REGPAT database was calculated as the share of backward citations to other patents originating from the same region.

$$Regionality = \frac{N\ regional\ citations}{N\ citations\ in\ REGPAT} \tag{4}$$

Impact

The number of citations patent received within 5 years after publication was used as a measure for impact. This indicator was directly adopted from the Patent Quality Indicator Database as made available by the OECD. This approach to quantify impact has been used by many others (i.e. Arts & Veugelers, 2013; Schoenmakers & Duysters, 2010; Squicciarini et al., 2013; Trajtenberg, 1990).

Because this data was strongly right-skewed, citation-counts+1 were log transformed.

Control variables

A number of control variables was obtained from various data sources. These control variables were introduced as covariates in the regression models where applicable. Prior to inclusion in the models, the control variables were tested for completeness; variables with large numbers of missing values were excluded. Subsequently, they were checked multicollinearity by calculating their Pearson correlations. In addition their variance inflation factor (VIF) using the vif() function of the USDM R package (Naimi, 2015). In the final models, all variables had VIF scores below 4, confirming the assumption of independence (Hair, Black, Babin, Anderson, & Tatham, 2006).

Table 1 Operationalisation of variables

Variable	Indicator	Calculation of scores/source	Measurement
		Dependent variables	
Radicality	Uniqueness of technology combinations	Average technological distance by jaccard index	Continuous
Impact	Citations within 5 years after publication	Number of citations	Count
		Independent variables	
Diversity evolutionary knowledge base	Diversity of technologies in backwards citations	Rao-Stirling diversity of technologies	Continuous
Regional knowledge diversity	Diversity of technologies in region	Rao-Stirling diversity of technologies	Continuous
Regionality	Extent of citations to patents in the same region	Number of regional citations / number of citations in REGPAT	Continuous
		Control variables	
Applicant type[A]	Type of applicant	Applicant type according to EEE-PPAT	Categorical
GDP per capita[BC]	Regional GDP per capita	OECD stats	Amount
Population density[C]	Population density (pop. per km2)	OECD stats	Count
Yearly new residents [BC]	Count of new residents in the region coming from another region of the country devided by region population	OECD stats	Count
Region type[B]	Regional typology from Urban(1) to Rural(5)	OECD stats	Nominal 1 to 5
Tertiary education[C]	Tertiary education (as % of labour force)	OECD stats	Percentage
R&D Expenditure[C]	R&D expenditure total (PPP)	OECD stats	Amount
R&D Personnel[C]	R&D personnel total (as % of employment)	OECD stats	Percentage

Note: Control variables on different aggregation levels: A = patent, B = TL3 level regions, C = TL2 level regions

*** In the end, the TL3 level variables (marked by B) could not be used due to large numbers of missing values. The variables were excluded from analysis .

The variable *applicant type,* as obtained from the EEE-PPAT database, was used to distinguish between corporate, academic, private and governmental applicants. This because universities are known to foster innovation through knowledge and human capital. Furthermore academic patents are expected to consider more radical innovations, as they are often based on breakthroughs in fundamental research (Schoenmakers & Duysters, 2010)

Finally a series of regional statistics was obtained from the OECS statistics database (OECD, 2015). These statistics were selected because they could to some extend measure the differences between regions the patents originate from, therewith possibly explaining part of the variations of the dependent variable. All variables from the OECD statistics database, as listed in Table 1, have been calculated as the mean value of all available data for the period 2003-2012. This, because not every statistic was available for every year for every region, hence a 10-year average comparison reduces the number of missing values. Most values were obtained as relative values, except for 'Yearly new residents' which was therefore divided by the population of the region. The completeness and level of detail of these data however varies, hence a limited number of factors were sufficiently available to include.

3.3 DATA ANALYSIS

By means of visualisations, descriptive statistics and regression analyses, the calculated and gathered variables were used to test the hypotheses formulated in the theoretical framework.

Technological proximity

First, in order to evaluate the validity of the technological proximity measure, and to gain a better understanding of this measure, the proximity values between all technology classes were used to perform a network analysis. First all 'edges' with a weight below 0.5 were omitted to create a better view. The Yifan Hu lay-out was used to arrange the nodes, and a modularity community detection algorithm was used to detect clusters and colour the nodes accordingly. Cluster labels were assigned manually.

Patent radicality

To determine whether the majority of innovations is indeed incremental, the distribution of patent radicality was evaluated. A density plot of patent radicality on a log scale with percentile indicators was created for this purpose. In addition, the skewness was calculated to measure the rate of skewedness.

Knowledge base diversity

The knowledge base of a patent, was operationalized as the set of backward citations of a single patent. It is assumed the patent is a product of knowledge spillovers between the technologies of its backward citations.

A scatterplot combined with a linear regression model was used to evaluate the presumed relation between knowledge base diversity and patent radicality. In the regression model, patent radicality was log transformed and the applicant type was used as control variable.

A density plot of the knowledge diversity was used to examine the natural maximum and majority of knowledge base diversity values.

Regionality

The degree of dependence of innovations on regional knowledge was expressed in a percentage of backward citations to patents from the same region.

Regional knowledge diversity

The final overarching hypothesis aimed at evaluating the supposed relation between regional knowledge diversity and patent radicality. First, scatterplots of region diversity versus patent radicality were drawn on both individual patent level and aggregated region level. These scatterplots were used to illustrate the relation between regional knowledge diversity and patent radicality.

Next, linear regression models were used to measure the strength and significance of the supposed relations. Each model was initially created with all relevant, independence checked control variables. This model was used as a starting point for a stepwise model selection with the stepAIC function of the MASS R package (Venables & Ripley, 2002). This algorithm searches for the best fitting model by iteratively adding and removing terms of the model until the best fit was obtained with the simplest model according to the Akaike Information Criterion (AIC) score. This model was then compared to the model with only control variables to check the contribution of the independent variabl.

Impact

The supposed high-risk/high-gain relationship between patent radicality and impact was tested by plotting the deviation in value distribution of impact for the full dataset and a subset of the most radical patents. The resulting graph shows which values of impact are over- and underrepresented in the subset of radical patents compared to the full dataset.

Bias effect test

Because knowledge diversity and patent radicality are both calculated from the technological proximity values between technology classes, the variables are partly based on the same data. This introduces an increased risk of bias, as the measured relation between the variables could be partially the result of the selected method. The significance of the bias can be quantified by measuring the impact of the technology classes of the radical patents on the knowledge diversity of its region.

For this purpose, a radicality threshold was estimated for the top 10% of patents in terms of radicality i.e. 305.8. The regional knowledge diversity was recalculated for a random sample of 100 regions, including only the technology classes assigned to the bottom 90% radicality patents.

Pierson correlations and comparison of regression models were used to determine the effect of the top-10% most radical patents on regional knowledge diversity.

4 RESULTS

In this section, the empirical findings of this study are presented. These findings will be used to critically assess each of the theoretical propositions regarding the relation between knowledge diversity and patent radicality. In addition, some additional results are discussed, starting with a technology map based on the technological proximity values.

4.1 TECHNOLOGICAL PROXIMITY

The first step of the data analysis was the calculation of relatedness values. These relative distances between all IFRIS technology classes were used for the majority of subsequent analyses, therefore the characteristics of these values were explored by evaluating the data distribution and mapping the technologies in a network. The distribution of relatedness values based on all PATSTAT applications up to 2012 is right-skewed (skewness: 69.05) and has a large share of zero values resulting from classes which never occurred together.

In order to grasp the meaning of the relatedness values, the technology classes and their proximity values were presented in a network visualisation (Figure 3).

The modularity clustered graph shows six dominant clusters of technology classes with material sciences at the core. This intuitive technology/knowledge network, where similar technologies are mapped closely together, suggests the technological proximity measure is valid.

A closer look into this network suggests that closely related technologies are often combined in inventions, because familiar technologies are located close to each other on this map.

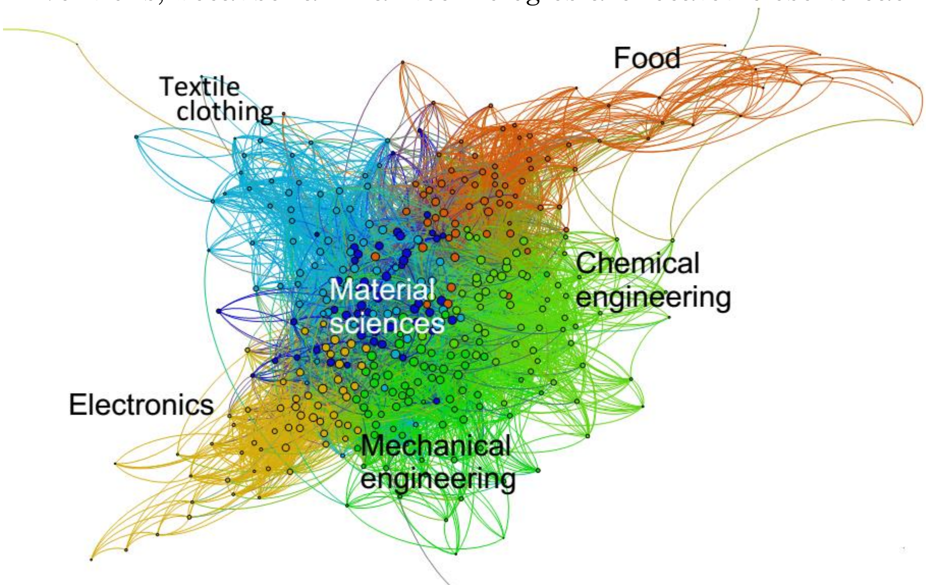


Figure 3 Network representation of technology classes in dataset

4.2 PATENT RADICALITY

From these technological relatedness values, the technological radicality was calculated for all priority patents filed between 1980 and 2012 with at least two technology classifications. These results were joined with the OECD REGPAT database. This resulted in a subset of 2,289,161 unique patent applications. Their radicality was strongly right-skewed, with a skewness of 80.93. After log normalisation, it is approximately normally distributed with a skewness of 0.193 and kurtosis of 3.20, as also visible in Figure 4. This confirms the particularly path dependent nature of innovation, as most inventions rely on closely related knowledge. Therefore I accept H1: The majority of innovations is incremental.

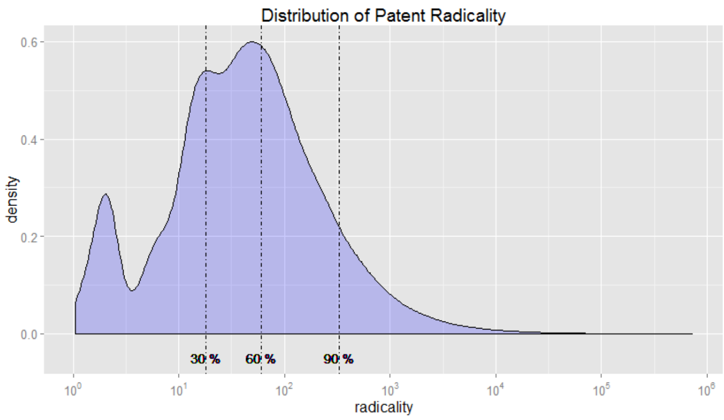


Figure 4 Distribution of Patent Radicality on logarithmic scale

In order to explore the outcome of this measure of *radicality,* a few highly radical applications were reviewed:

Table 2 listing of most radical patents

Application	Title	Ifris class
328793888	Process for structuring clothing	T03F34,Corsets
EP20100191499		T01F06,Mechanic Digital Comput
15798122	Piezoelectrically controlled active wear	T01F08,Semiconductor Devices
EP20000988154		T03F34,Corsets
56262543	a mobile concentration system and method for milk	T01F32,Vehicles
EP20080723993		T04F18,Dairy Products

These examples all exhibit unique combinations of IFRIS classes which intuitively do not seem to be conventional, such as 'corsets' and 'semiconductor devices'.

4.3 KNOWLEDGE BASE DIVERSITY

The knowledge base diversity of an innovation, as based on the set of backwards citations of a patent, was expected to be positively related with patent radicality. It was found that the log-transformed radicality has a moderate positive, highly significant ($p<0.001$) correlation with the diversity of its knowledge base. Based on the regression model in Table 3 and the scatterplot in Figure 5 , I accept H2 because it is clear that more radical innovations stem from more diverse knowledge sources than more incremental innovations. The regression model with an R^2 of 0.137 shows that the variation in patent radicality can be attributed for an important share to the diversity of a patent's knowledge base. The control model, with an R^2 of 0.017 confirms that the effect is not the product of the control variables. This illustrates the diversity of knowledge in an innovations evolutionary knowledge base is reflected in diversity of the core-technologies of an innovation.

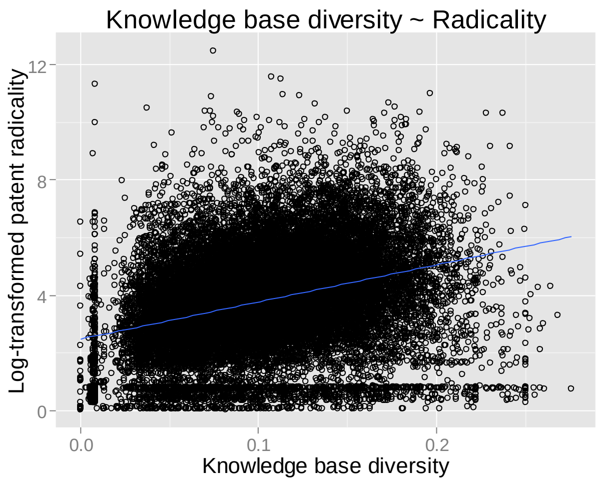


Figure 5 Patent radicality versus knowledge base diversity

Table 3 regression models of radicality and knowledge base diversity

	Dependent variable	
	Radicality (log-transformed)	
	control	**model**
Knowledge base diversity		12.741*** (0.201)
applicant_type GOV NON-PROFIT	-0.363*** (0.103)	-0.288*** (0.096)
applicant_type HOSPITAL	-1.226*** (0.430)	-0.884** (0.403)
applicant_type INDIVIDUAL	0.566*** (0.040)	0.464*** (0.037)
applicant_type UNIVERSITY	-0.727*** (0.096)	-0.633*** (0.090)
applicant_type MULTI	-0.241*** (0.021)	-0.252*** (0.019)
Constant	3.834*** (0.014)	2.579*** (0.024)
Observations	28,874	28,874
Adjusted R^2	0.017	**0.137**
Residual Std. Error	1.663 (df = 28868)	1.559 (df = 28867)
F Statistic	101.440*** (df = 5; 28868)	762.762*** (df = 6; 28867)
Note:		*p<0.1; **p<0.05; ***p<0.01

Furthermore, as Figure 6 indicates, the knowledge base diversity is nicely distributed (skewness=0.225, kurtosis=2.83) around a mean diversity of 0.094 with a maximum of 0.278, and a large majority below 0.18, which is the lowest occurring regional knowledge diversity. Hence, it can be concluded that the natural maximum diversity of knowledge sources combined into new patents is well below the theoretical maximum of 1, and the vast majority of patents rely on a low diversity of knowledge sources. Therefore I accept H4 and H5: The diversity of knowledge sources combined into innovations has a 'natural' maximum well below the theoretical maximum and the majority of innovations stems from technologically similar knowledge.

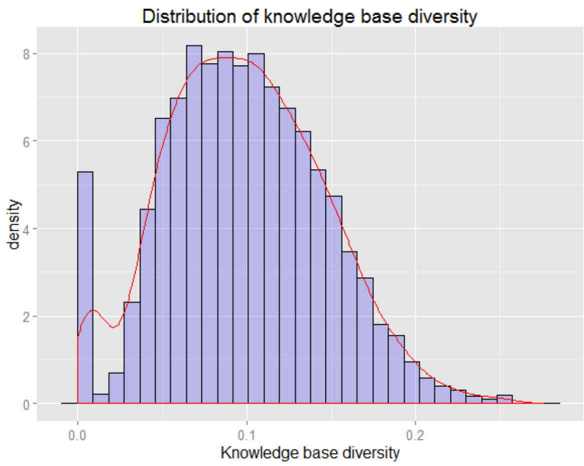


Figure 6 distribution of knowledge base diversity

4.4 Regionality

Having confirmed the technological rootedness or path dependency of innovation, the degree of geographical rootedness or place dependency is the next step towards answering the research question.

A simple summary of the regionality data (Table 4) shows that on average patents rely for 27% on regional knowledge, and the top 25% of the patents in terms of regionality rely for at least 50% on regional knowledge. These values represent a significant share of the knowledge sources of a patent, indicating knowledge production is to an important degree place dependent. H6 is therefore accepted: Innovations rely for a significant share on regional knowledge.

Table 4 Summary of patent regionality

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.000	0.000	0.000	0.238	0.500	1.000

4.5 REGIONAL KNOWLEDGE DIVERSITY

While the former results have confirmed the theorized mechanisms underlying the presumed relation between regional knowledge diversity and patent radicality, the correlation between these factors was also empirically tested.

Initial linear regressions between region diversity (TL2, TL3) and patent radicality showed a highly significant ($p < 2.2e^{-16}$) correlations with negligible effect ($R^2 = 0.0005$ and 0.001, see Appendix I for regression table). A scatterplot of these variables (Figure 7) however suggests a relation between region diversity and patent radicality exists. A high region diversity seems to have an *enabling* effect on patent radicality. Yet, a high region diversity does not automatically result in highly diverse patents.

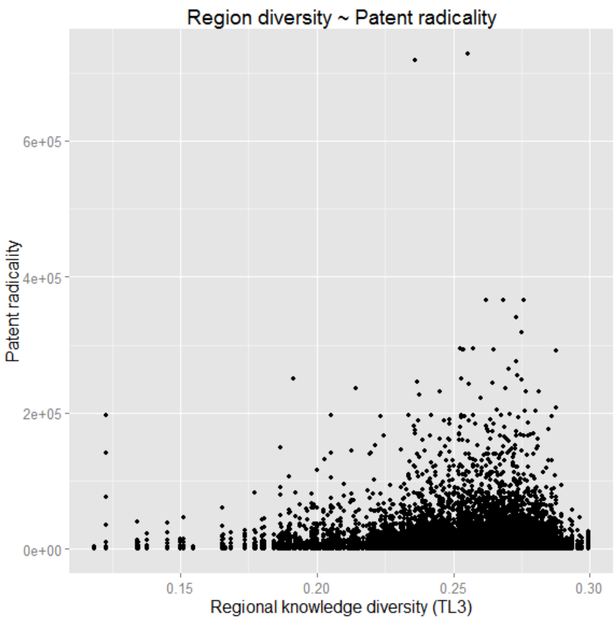


Figure 7 Patent radicality versus region diversity

When averaged per region, in order to reduce some of the noise, the relation between patent radicality and regional knowledge diversity becomes both stronger and more clear (Figure 8). The scatterplot of these variables shows a positive correlation and the results of the linear regression model, as listed in Table 5, indicate a highly significant effect: The model on TL2 and TL3 level have R^2 values of 0.286 and 0.166 repectively. While the control variables account for an important part for the variation, regional knowledge diversity still has a relevant effect. This confirms that a part of the variation in average patent radicality can be explained by variations in regional knowledge diversity. The regression also strongly implies a significant effect of applicant type, number of patents and education level of workforce. Applicant types 'Individual' and 'Company' both have a positive effect on patent radicality, whereas applicant types 'Gov Non-profit' and 'University' as well as 'N patents' and 'Education level workforce' have a negative influence on patent radicality. These findings will be addressed in the discussion.

So while the explanatory power of regional knowledge diversity on patent radicality is limited to an 'enabling effect' on individual patent level, it is much stronger and highly significant on region level. Furthermore, the majority of the underlying mechanisms, as argued in the theoretical framework have been confirmed by empirical evidence by testing the former hypotheses. Therefore, H7 is accepted: Highly radical innovations are more likely to emerge from regions with greater knowledge diversity.

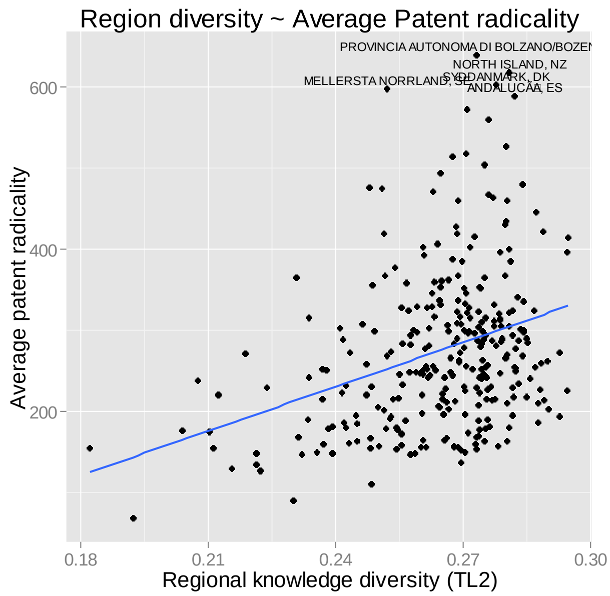


Figure 8 Average patent radicality per region versus regional knowledge diversity

Table 5 Regression analysis of region averaged patent radicality

	Dependent variable:			
	Average Radicality			
	control TL2	**model TL2**	**control TL3**	**model TL3**
Regional knowledge diversity		1,670.676***		1,123.976***
N patents	-0.001**	-0.001***	-0.003***	-0.003***
Education level workforce	-1.392	-1.715**		
GDP per capita	0.001	0.001		
Population density	-0.016**	-0.011		
Applicant GOV NON-PROFIT	-399.147	-563.925*		
Applicant UNIVERSITY	-802.662***	-854.756***		
Applicant COMPANY			230.982***	245.957***
Applicant INDIVIDUAL	751.151***	541.059***	863.083***	687.968***
Observations	257	257	1,101	1,101
R^2	0.234	0.309	0.123	0.169
Adjusted R^2	0.213	**0.286**	0.120	**0.166**
Residual Std. Error	86.481 (df = 249)	82.334 (df = 248)	117.312 (df = 1097)	114.238 (df = 1096)
F Statistic	10.880*** (df = 7; 249)	13.843*** (df = 8; 248)	51.077*** (df = 3; 1097)	55.604*** (df = 4; 1096)
Note:				*p<0.1; **p<0.05; ***p<0.01

Bias effect test

In order to exclude the possibility that the positive effect is a result of a methodological bias, the regional diversity was recalculated without the technology classes of the 10 % most radical patents. A pearson correlation of the regional knowledge diversity in the original dataset and the corrected dataset returns a highly significant ($p < 2.2e^{-16}$) correlation of 0.9990908, indicating both datasets measure effectively the same. In addition, their predictive power on radicality was compared. These results (Table 6) also confirm patent radicality is not just predicted by regional knowledge diversity because the variables are partially based on the same data.

Table 6 Comparison of the effect of region diversity and corrected region diversity on patent radicality to rule out bias.

	Dependent variable:	
	Average Radicality	
	corrected model	original model
Regional knowledge diversity	1,386.260*** (360.353)	1,399.146*** (371.757)
Constant	-92.542 (91.019)	-98.316 (94.538)
Observations	99	99
R^2	0.132	0.127
Adjusted R^2	0.123	0.118
Residual Std. Error (df = 97)	101.627	101.916
F Statistic (df = 1; 97)	14.799***	14.165***
Note:	*p<0.1; **p<0.05; ***p<0.01	

4.6 Impact

Thus far, the theoretical insights regarding the origins of technologically radical innovations have been confirmed. The final proposition was regarding the high-risk/high-gain attribute of radical innovations, as measured in impact. The overall frequency distribution of impact is highly right-skewed (skewness=9.36) as most patents received little to no citations at all. This is true for the full dataset as well as for the most radical patents, as is shown in Figure 9.

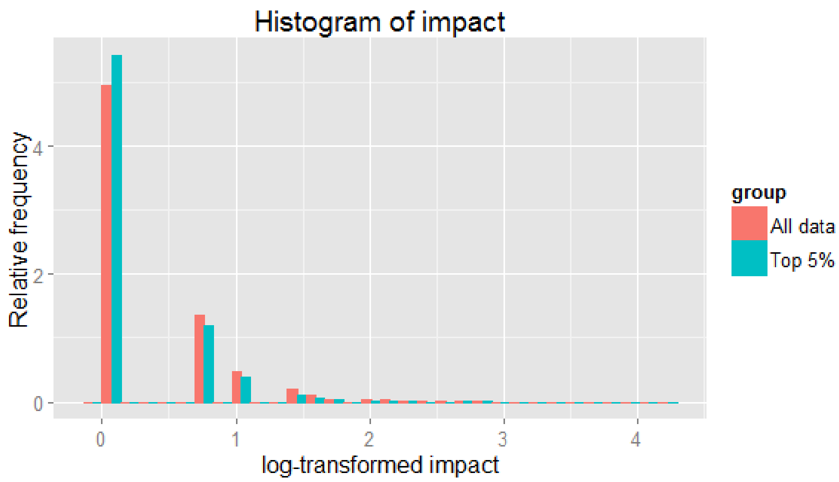


Figure 9 Histogram of relative frequency patent impact

More interesting are differences in the relative frequencies of impact between the full dataset and the most radical patents, as shown in Figure 10. In this figure the impact distribution of the 5% most radical patents (radicality >= 952.2) is compared to the impact distribution of all patents. These results indicate that low-impact patents are over-represented amongst the most radical patents, while the impact in the mid-range is under-represented. In the high-impact range, there are no significant differences between the most radical patents and the average. These results confirm the high-risk attribute, yet there is no evidence of high-gain. H3 is therefore rejected, radical innovations do not have a high-risk high/gain characteristics in terms of impact.

Since the presumed relation between patent radicality and impact could not be confirmed, other analyses were performed on the data, to further explore the dynamics of these variables. The scatter plot of impact and log-normalized radicality (Figure 11) implicates a structured relation between impact and patent radicality. The seems to be related to patent radicality in the way that radicality is both an enabling and limiting force on impact. A regression of a quadratic model[5] shows that the large number of zero-cited patents weighs down the least-squares fit, resulting in a model with an adjusted R-squared of 0.0012 which barely explains the variation. The graph in Figure 11 however, suggests that a certain amount of radicality is beneficial for a patents impact, yet highly radical patents fail to gain a foothold.

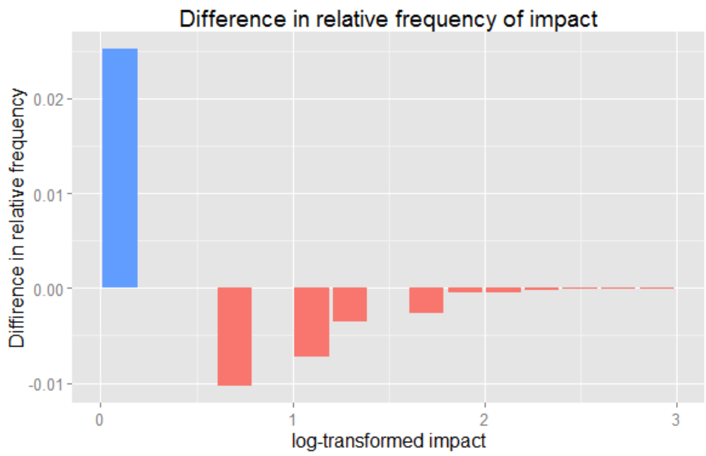


Figure 10 frequency distribution of 5% most radical patents, in comparison to the frequency distribution of the full dataset.

Rather than confirming the high-risk/high-gain relation, these results implicate that the influence of raciality on impact is in line with the model of optimal cognitive distance and absorptive capacity, where innovative performance is expressed in an inverted-U shape (Nooteboom et al., 2007).

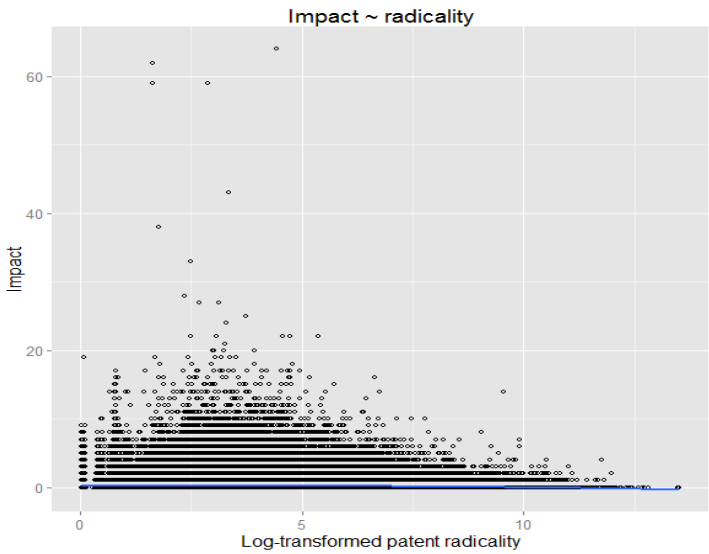


Figure 11 Scatter of the patent impact versus patent radicality with quadratic regression line (blue)

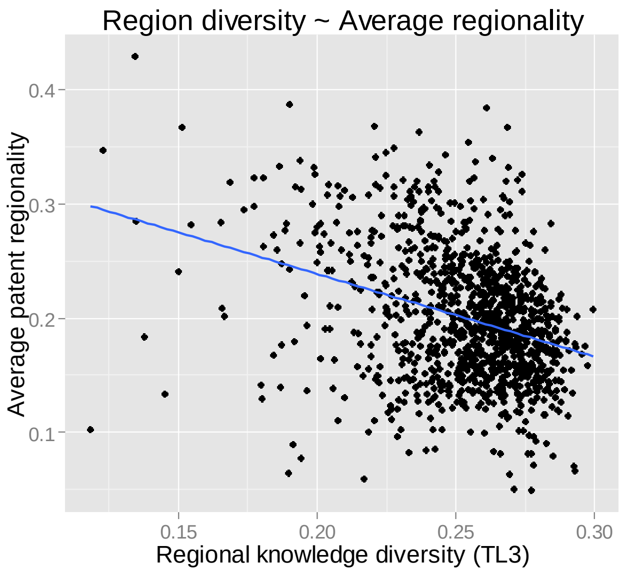
[5] Full regression table of Impact and Radicality can be found in Appendix I

4.7 OTHER RESULTS

In addition to the earlier discussed results, there were results that were not hypothesized, but still interesting in the context of this research.

Region diversity and average regionality

The relation between regional knowledge diversity and average radicality per region is such a result. A scatterplot of these variables clearly shows a negative relation, indicating that regions with more diverse knowledge rely more on knowledge sources from other regions. This implies that if it is the goal to diversify the regional knowledge base, inter-regional knowledge flows should be promoted. Further research into this matter would be required to be able to explain the forces, mechanisms and implications of reduced regionality, and the nature of its relation with regional knowledge diversity.



Geographic projections of data

The advantage of the REGPAT database is the availability of cleaned address data and region data. These data were used to project the variables on maps of Europe, the United States and on a globe. While outside the scope of this research the graphs show for example that knowledge diversity of regions is quite evenly distributed across Europe, while the number of patents per region is concentrated in a small number of regions. Additional visualisations of region level data are included in Appendix III, projections on an interactive world globe can be accessed using a Chrome browser on http://www.hutschemaekers.com/globe.

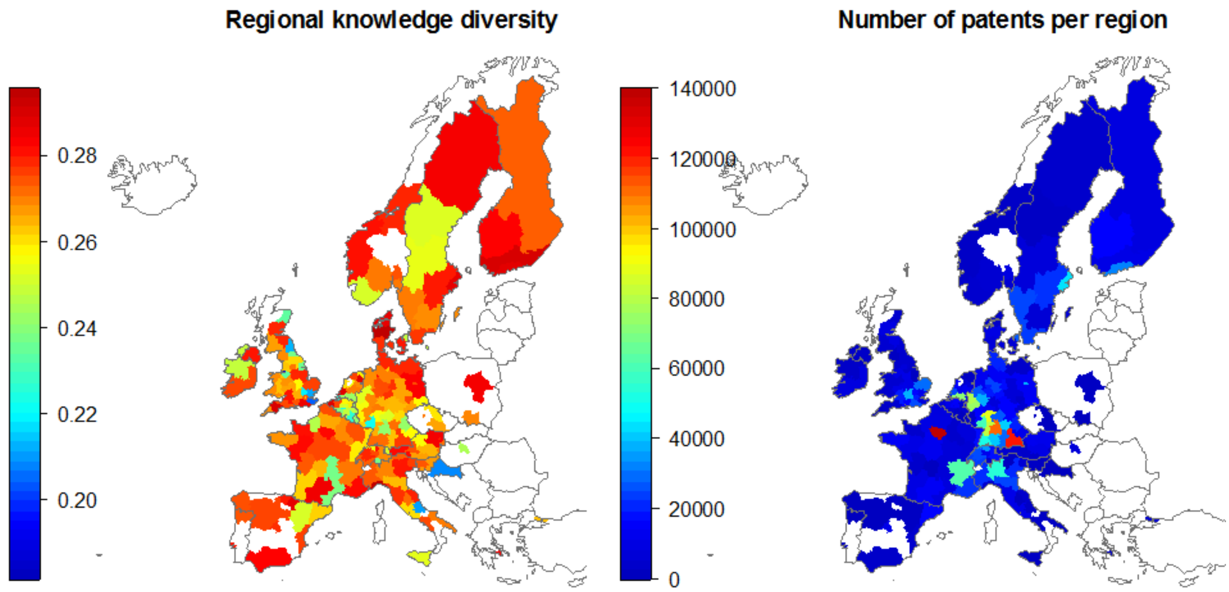


Figure 12 Knowledge diversity and patent counts of European TL2 regions

5 DISCUSSION

The objective of this research was to gain insights in the technological and geographical origins of radical innovation. The theoretical framework, predicted that regions with a diverse knowledge base would be a fertile environment for knowledge spillovers between unrelated pools of knowledge, resulting in increased potential for radical innovation. In addition it was argued, that radical innovations, while facing an increased risk of failure, would have an increased potential of success. Although the framework was partially based on empirically supported claims, empirical evidence regarding the technological and geographical origins of radical innovation has been lacking thus far (Arts & Veugelers, 2013; Schoenmakers & Duysters, 2010).

The present study has provided evidence, based millions of patents, from thousands of regions across OECD countries, regardless of their field of origin and language, that the conceptualized mechanisms and forces of the radical innovation process can be confirmed. The high-risk/high-gain conjecture on the contrary, was not confirmed: Only the increased risk of radical innovations was revealed, while no trace of the increased potential was found. Within the limitations of this research, as mostly incurred by the methodology and data selection, these results have a number of theoretical and societal implications.

5.1 THEORETICAL IMPLICATIONS

The results of this study confirm the positive influence of technological and geographical proximity on innovation, which is in line with earlier findings in literature, stressing the importance of these proximities for knowledge spillovers (Jaffe, 1989; Marrocu et al., 2014). Furthermore, in line with the theoretical framework, examples of highly radical innovation show that radical innovations are indeed recombinations of previously existing knowledge and that increased technological distance between knowledge sources leads to more radical innovation (Boschma, 2005; Nooteboom et al., 2007). Although the effects of knowledge base diversity and regional knowledge diversity on patent radicality were evident, the strong noise in the data implicates many other factors play a role in the radical innovation process. This finding confirms the 'confusion' regarding the origins of radical innovation, as suggested by Arts and Veugelers (2013), and suggests that the mechanisms of radical innovation are unlikely to be captured in a single clear model. The regression results do indicate a significant effect of applicant type on patent radicality. Applicant type 'individual' has the strongest positive effect, followed by 'company' while types 'hospital', 'university' and 'government non-profit' have a negative effect. This finding could implicate market entrants indeed play a dominant role in radical innovation (C. Christensen & Rosenbloom, 1995; Henderson, 1993). Moreover this finding suggests that the products of scientific research i.e. patents applied by universities, research institutes and hospitals tend to be incremental innovations. This finding aligns with Kuhn's (1962) notion of incommensurability and scientific paradigms, stating that the majority of scientific developments are accumulations within a paradigm. Yet both implications inferred from the effect of the control variables need further research, because they are outside the theoretical scope of this study and other explanations for the effects cannot be ruled out.

The missing evidence of enhanced high-gain potential of radical innovations confirms that radical innovations often have difficulties to gain foothold. This implies the supposed pay-off of radical innovation is missing or clearly not as strong as expected. The findings support the theory of optimal cognitive distance (Nooteboom et al., 2007), and the concepts of technology space (Hidalgo, Klinger, Barabási, & Hausmann, 2007), both indicating that the greatest innovation potential is within a limited distance of the existing base of knowledge and capabilities.

5.2 Societal implications

Results indicate that diversified regions have increased potential for radical innovation, therefore diversification would be a suitable strategy for promoting radical innovation in a region. Moreover, the analysis of the relation between regionality and region diversity suggest promoting inter-regional knowledge spillovers is beneficial for the knowledge diversity in a region, and therefore radical innovation. However, results also indicate that radical innovations may not have the increased potential for success as suggested in theory. Moreover, the findings suggest that high impact is not associated with radical innovation, but most likely achieved by innovations incorporating moderately diverse knowledge.

This implicates that innovations can be too radical and should have a certain degree of compatibility with established standards, in order to gain sufficient momentum to take off. Furthermore, it seems that in order to maximise innovation potential in terms of impact, regional policy should stimulate keeping a diverse knowledge base within a limited range of cognitive distance, as previously suggested on country level by Hidalgo et al. (2007). Regional expansion strategies should focus on acquiring knowledge capital within a limited technological proximity of their current knowledge base.

5.3 QUALITY AND LIMITATIONS

The selection of data and methodology poses some implications for the reliability and validity of the research.

This research relies on publicly available patent data, and uses publicly available methods for data processing and analysis. Furthermore, the routines, queries and scripts used in the research are available on GitHub[6]. Therefore, the study could easily replicated or extended by any researcher who desires to do so.

The use of patent data from all OECD countries, regardless of industry and language, enhances the generalizability of the results. This wealthy source of data however, also introduces some limitations: Patents describe inventions in technologies, applications and processes, for which legal protection is sought. As such, they do not represent a one-on-one translation of innovation (Heimeriks, Alkemade, Schoen, Villard, & Laurens, n.d.). In addition, not all inventions are patented, and the REGPAT database only entails patent applications filed through EPO or PCT with address information available. This restriction to EPO and PCT patent applications increases comparability of the individual applications, yet limits the sample to inventions for which transnational protection is sought. In addition, the technology classes used are assigned to patents for administrative purposes, not to reflect the knowledge constituting a patent. Finally patenting behaviour is dependent on industry, country and patenting institution, and time, which should be considered carefully to prevent unwittingly biasing (De Rassenfosse et al., 2013; Heimeriks et al., n.d.). Nevertheless, patent data offers an unparalleled wealth of rich, structured data on novelty production, therefore the drawbacks and limitations are to be accepted.

In addition, some methodology-specific limitations apply: The regions used as a measure for geographical proximity, the TL2 and TL3 regions as defined by the OECD, have limited comparability. Their size in terms of surface or population, the degree of administrative independence, and many other parameters very vastly within this dataset. (E.g. a TL 2 region in the Netherlands is a province, while a TL2 region in the US is a state.) The use of control variables partially compensates for this problem, but on many parameters, insufficient data was available to control for the differences.

Moreover, the approach of radicality used in the present study, is very likely to produce a left-skewed distribution of radicality. This means that with any real dataset, H1 is very

[6] The repository with the scripts used in this thesis can be accessed through https://github.com/Guus-H/thesis

unlikely to be rejected. Yet given the wide consensus on this particular aspect of radical innovation, this does not pose a big problem.

Finally, despite the limitations, this research provides new empirical evidence for important theories regarding the origins of innovation, on very large scale, based on millions of patents applied around the world, in all technology classes.

5.4 FUTURE RESEARCH

The lack of high-gain potential of radical innovation raises questions worth investigating. In particular if the high gain attribute is truly absent for radical innovations, why this is, and if within a particular range of cognitive distance the effect still applies.

Other questions arose from the control variables. As mentioned, the strong and significant effect of the control variable 'applicant type', suggest market entrants may have a dominant role over established firms in radical innovation, and scientific research occurs mostly within established paradigms. Yet these findings were not theorized, and alternative explanations for these findings have not been researched. The findings to raise questions, and further research between applicant type, and innovation radicality could provide valuable insights on the sources of radical innovation.

Besides the applicant type, both covariates 'N patents' and 'Education level workforce' have significant negative influence on average patent radicality in a region. These findings intuitively seem odd, and no straightforward explanation is obvious. Further exploration of these dynamics would be required to say something meaningful about these figures.

6 CONCLUSION

Based on existing theory, I propose a model where radical innovations are recombinations of existing knowledge, which are conceived through knowledge spillovers between technologically diverse knowledge sources. These knowledge sources with high technological distance between them are more likely to be combined, if they can benefit from the advantages of high geographical proximity for knowledge spillovers. Hence, regions with a diverse knowledge portfolio are believed to be a fruitful environment for radical innovation. Whereas this model is theoretically supported, empirical evidence is scarce.

This led to the following research question:

Does regional knowledge diversity stimulate radical innovation?

Results confirm radical innovations are indeed the result of recombinations of existing knowledge. Geographical and technological proximity of the knowledge sources both have considerable positive influence on the innovation process.

Moreover, the results affirm that combinations of more diverse sources of knowledge lead to more radical innovations. In addition, it was confirmed that high geographical proximity operationalized as regions, can function as a fruitful environment for the radical innovation process.

Finally the results confirmed that radical innovations indeed have an increased risk of failure, many technological radical innovations fail to take off, yet the presumed compensation, the high-gain property, lacked completely. This finding raises questions regarding the supposed benefits of innovating radically, for both the inventor and society.

7 References

Abernathy, W. J., & Utterback, J. M. (1978). Patterns of Industrial Innovation. *Technology Review*, *80*(7), 40–end.

Ahuja, G., & Lampert, C. M. (2001). Entrepreneurship in the large corporation: A longitudinal study of how established firms create breakthrough inventions. *Strategic Management Journal*, *22*(6-7), 521–543. doi:10.1002/smj.176

Aldieri, L. (2011). Technological and geographical proximity effects on knowledge spillovers: evidence from the US patent citations. *Economics of Innovation and New Technology*, *20*(6), 597–607. doi:10.1080/10438599.2011.554632

Aldieri, L. (2013). Knowledge technological proximity: evidence from US and European patents. *Economics of Innovation and New Technology*, *22*(8), 807–819. doi:10.1080/10438599.2013.788838

Amin, A., & Wilkinson, F. (1999). Learning, proximity and industrial performance: an introduction. *Cambridge Journal of Economics*, 121–125.

Andersen, N. (2014). Behavioral traps and innovation. *Deloitte Review*, (15), 44–55.

Arts, S. (2012). Path dependency and recombinant exploration: How established firms can outperform in the creation of technological breakthroughs. *Katholieke Universiteit Leuven Department of Managerial Economics, Strategy and Innovation Working Paper*, (1228), 1–37.

Arts, S., & Veugelers, R. (2013). The Technological Origins and Novelty of Breakthrough Inventions. *SSRN Electronic Journal*. doi:10.2139/ssrn.2230366

Baregheh, A., Rowley, J., & Sambrook, S. (2009). Towards a multidisciplinary definition of innovation. *Management Decision*, *47*(8), 1323–1339. doi:10.1108/00251740910984578

Boschma, R. (2005). Proximity and Innovation: A Critical Assessment. *Regional Studies*, *39*(1), 61–74. doi:10.1080/0034340052000320887

Boschma, R., Heimeriks, G., & Balland, P.-A. (2014). Scientific knowledge dynamics and relatedness in biotech cities. *Research Policy*, *43*(1), 107–114. doi:10.1016/j.respol.2013.07.009

Carlino, G. (2001). Knowledge spillovers: cities' role in the new economy. *Business Review*, (2000), 17–26.

Chandy, R., & Tellis, G. (2000). The incumbent's curse? Incumbency, size, and radical product innovation. *Journal of Marketing*, *64*(3), 1–17.

Christensen, C. M., & Raynor, M. E. (2003). *The Innovator's Solution*. Harvard Business Review Press.

Christensen, C., & Rosenbloom, R. (1995). Explaining the attacker's advantage: Technological paradigms, organizational dynamics, and the value network. *Research Policy*, *24*, 233–257.

CIB. (2009). Corporate Invention Board: Sources of the analyzed data. Retrieved March 6, 2015, from http://www.corporateinventionboard.eu/en/data-sources

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

Dahlin, K. B., & Behrens, D. M. (2005). When is an invention really radical? *Research Policy*, *34*(5), 717–737. doi:10.1016/j.respol.2005.03.009

De Rassenfosse, G., Dernis, H., Guellec, D., Picci, L., & Van Pottelsberghe De La Potterie, B. (2013). The worldwide count of priority patents: A new indicator of inventive activity. *Research Policy*, *42*(3), 720–737. doi:10.1016/j.respol.2012.11.002

Fleming, L. (2001). Recombinant uncertainty in technological search. *Management Science*, *47*(1), 117–132.

Freeman, C. (1992). The Economics of Hope: Essays on Technical Change and Economic Growth. *Pinter, London*.

Garcia, R., & Calantone, R. (2002). A critical look at technological innovation typology and innovativeness terminology: A literature review. *Journal of Product Innovation Management*. doi:10.1016/S0737-6782(01)00132-1

Gerken, J. M., & Moehrle, M. G. (2012). A new instrument for technology monitoring: novelty in patents measured by semantic patent analysis. *Scientometrics*, *91*(3), 645–670. doi:10.1007/s11192-012-0635-7

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate data analysis* (Vol. 6). Pearson Prentice Hall Upper Saddle River, NJ.

Hargadon, A. (2003). *How Breakthroughs Happen: The Surprising Truth about how Companies Innovate*. Harvard Business Press.

Heimeriks, G., Alkemade, F., Schoen, A., Villard, L., & Laurens, P. (n.d.). National and Sectoral characteristics of the internationalization of the inventive activity of multinational corporations.

Heimeriks, G., Alkemade, F., Schoen, A., Villard, L., & Laurens, P. (2013). *The Evolution of Technological Knowledge: A co-evolutionary analysis of patterns of Corporate Invention. 35th DRUID Celebration Conference*.

Heimeriks, G., & Boschma, R. (2013). The path- and place-dependent nature of scientific knowledge production in biotech 1986-2008. *Journal of Economic Geography*, *14*(2), 339–364. doi:10.1093/jeg/lbs052

Henderson, R. (1993). Underinvestment and incompetence as responses to radical innovation: Evidence from the photolithographic alignment equipment industry. *The RAND Journal* *of Economics*.

Hidalgo, C. A. a, Klinger, B., Barabási, A. a-L., & Hausmann, R. (2007). The product space conditions the development of nations. *Science, 317*(July), 482–488. doi:10.1126/science.1144581

Iammarino, S., & McCann, P. (2006). The structure and evolution of industrial clusters: Transactions, technology and knowledge spillovers. *Research Policy*, *35*(7), 1018–1036. doi:10.1016/j.respol.2006.05.004

Jaffe, A. B. (1986). Technological opportunity and spillovers of R&D: evidence from firms' patents, profits and market value, *76*(5), 984–1001.

Jaffe, A. B. (1989). Real effects of academic research. *The American Economic Review*, *79*(5), 957–970.

Jaffe, A. B., Trajtenberg, M., & Henderson, R. (1993). Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *The Quarterly Journal of Economics*, *108*, 577–598. doi:10.2307/2118401

Keijl, S. (2011). *Innovation Radicalness: Exploring the Relationship between Novelty and Impact.*

Keizer, J. a., & Halman, J. I. M. (2009). Risks in major innovation projects, a multiple case study within a world's leading company in the fast moving consumer goods. *International Journal of Technology Management*, *48*(4), 499. doi:10.1504/IJTM.2009.026691

Knoben, J., & Oerlemans, L. a. G. (2006). Proximity and inter-organizational collaboration: A literature review. *International Journal of Management Reviews*, *8*(2), 71–89. doi:10.1111/j.1468-2370.2006.00121.x

Kogler, D. F., Rigby, D. L., & Tucker, I. (2013). Mapping Knowledge Space and Technological Relatedness in US Cities. *European Planning Studies*, *21*(9), 1374–1391. doi:10.1080/09654313.2012.755832

Kuhn, T. S. (1962). The Structure of Scientific Revolutions. *International Encyclopedia of Unified Science*, *II*(2), 92–210. doi:10.1111/j.1468-0149.1963.tb00813.x

Leydesdorff, L., Alkemade, F., Heimeriks, G., & Hoekstra, R. (2015). Patents as instruments for exploring innovation dynamics: geographic and technological perspectives on "photovoltaic cells." *Scientometrics*, *102*(1), 629–651. doi:10.1007/s11192-014-1447-8

Leydesdorff, L., Kushnir, D., & Rafols, I. (2014). Interactive overlay maps for US patent (USPTO) data based on International Patent Classification (IPC). *Scientometrics*, *98*(3), 1583–1599. doi:10.1007/s11192-012-0923-2

Lychagin, S., Pinkse, J., Slade, M., & Reenen, J. Van. (2010). Spillovers in space: does geography matter? *National Bureau of Economic Research*, (991).

Mackinnon, D., Cumbers, A., & Chapman, K. (2002). Learning, innovation and regional development: a critical appraisal of recent debates. *Progress in Human Geography*, *26*(3), 293–311. doi:10.1191/0309132502ph371ra

Marrocu, E., Paci, R., & Usai, S. (2014). The complementary effects of proximity dimensions on knowledge spillovers. *Spatial Economic Analysis*, *9*(1), 9–30.

Maurseth, P., & Verspagen, B. (2002). Knowledge spillovers in Europe: a patent citations analysis. *The Scandinavian Journal of …*, *104*(4), 531–545.

Naimi, B. (2015). usdm: Uncertainty Analysis for Species Distribution Models.

Neffke, F. (2009). *Productive Places: The influence of technological change and relatedness on agglomeration externalities*. Utrecht University.

Nooteboom, B., Van Haverbeke, W., Duysters, G., Gilsing, V., & van den Oord, A. (2007). Optimal cognitive distance and absorptive capacity. *Research Policy*, *36*(7), 1016–1034. doi:10.1016/j.respol.2007.04.003

OECD. (2015). Demographic Statistics. Retrieved January 29, 2015, from http://stats.oecd.org/Index.aspx?datasetcode=REG_DEMO_TL2

Oerlemans, L., Meeus, M., & Boekema, F. (2001). On the spatial embeddedness of innovation networks: An exploration of the proximity effect. *Tijdschrift Voor Economische En Sociale Geografie*, *92*(1), 60–75. doi:10.1111/1467-9663.00139

Ponds, R., van Oort, F., & Frenken, K. (2007). The geographical and institutional proximity of research collaboration. *Papers in Regional Science*, *86*(3), 423–443. doi:10.1111/j.1435-5957.2007.00126.x

R Core team. (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rosenberg, N. (1994). *Exploring the black box: Technology, economics, and history*. Cambridge University Press.

RStudio Team. (2015). RStudio: Integrated Development Environment for R. Boston, MA.

Schilling, M. A., & Green, E. (2011). Recombinant search and breakthrough idea generation: An analysis of high impact papers in the social sciences. *Research Policy*, *40*(10), 1321–1331.

Schoen, A., Villard, L., & Larédo, P. (2014). Corporate Invention Board (CIB) Database. IFRIS.

Schoenmakers, W., & Duysters, G. (2010). The technological origins of radical inventions. *Research Policy*, *39*(8), 1051–1059. doi:10.1016/j.respol.2010.05.013

Schumpeter, J. A. (1942). Creative destruction. In *Capitalism, socialism and democracy* (p. 1975). Harper, New York.

Singh, J., & Fleming, L. (2010). Lone Inventors as Sources of Breakthroughs: Myth or Reality?

Management Science, *56*(1), 41–56. doi:10.1287/mnsc.1090.1072

Squicciarini, M., Dernis, H., & Criscuolo, C. (2013). Measuring Patent Quality: Indicators of Technological and Economic Value. *OECD Science, Technology and Industry Working Papers*, 70. doi:10.1787/5k4522wkw1r8-en

Stirling, A. (2007). A general framework for analysing diversity in science, technology and society. *Journal of the Royal Society, Interface / the Royal Society*, *4*(February), 707–719. doi:10.1098/rsif.2007.0213

Thoma, G., & Torrisi, S. (2007). *Creating powerful indicators for innovation studies with approximate matching algorithms. A test based on PATSTAT and Amadeus databases*. *Conference on Patent Statistics for Policy Decision Making*.

Thoma, G., Torrisi, S., Gambardella, A., & Guellec, D. (2010). Harmonizing and combining large datasets–an application to firm-level patent and accounting data.

Tidd, J., Bessant, J., & Pavitt, K. (2005). *Managing Innovation* (Third.). Wiley.

Utterback, J. (1996). *Mastering the Dynamics of Innovation*. Harvard Business Press.

Uzzi, B., Mukherjee, S., Stringer, M., & Jones, B. (2013). Atypical combinations and scientific impact. *Science (New York, N.Y.)*, *342*, 468–72. doi:10.1126/science.1240474

Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S* (Fourth.). New York: Springer.

Verspagen, B. (2006). Measuring Intersectoral Technology Spillovers: Estimates from the European and US Patent Office Databases. *Economic Systems Research*, *9*(1), 47–65. doi:10.1080/09535319700000004

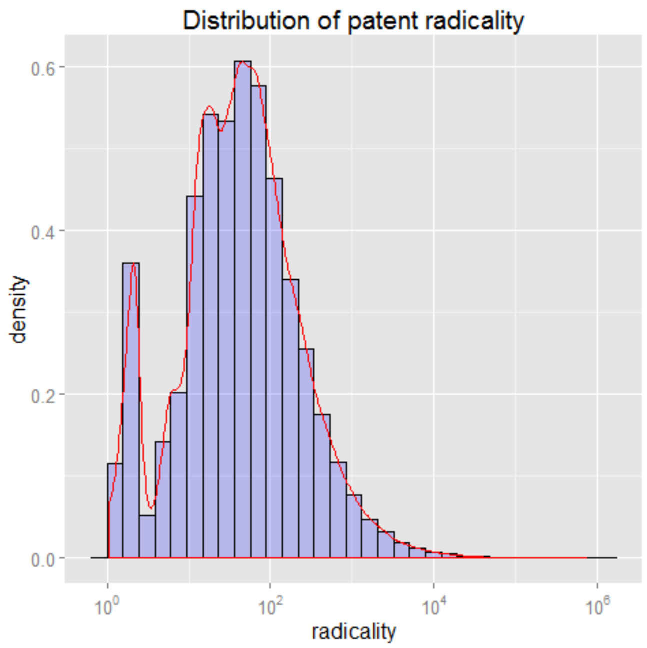
Wickham, H. (2015). bigrquery: An Interface to Google’s BigQuery API.

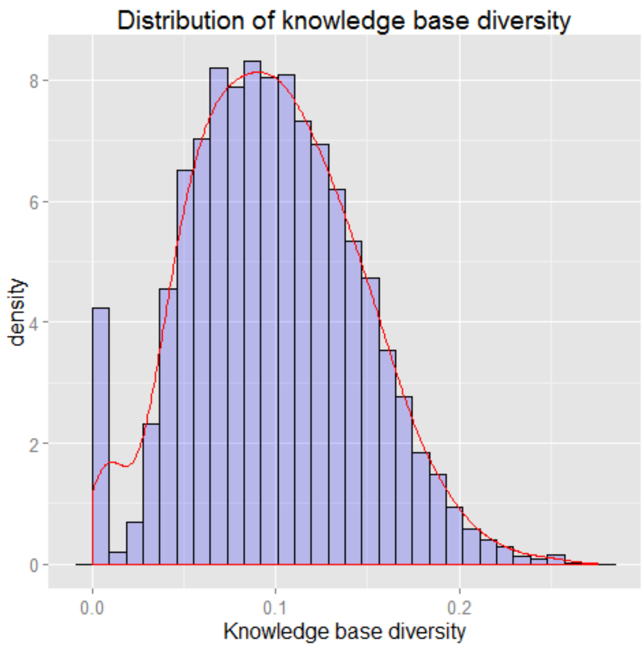
Winter, S. G., & Nelson, R. R. (1982). *An Evolutionary Theory of Economic Change*. Harvard University Press.

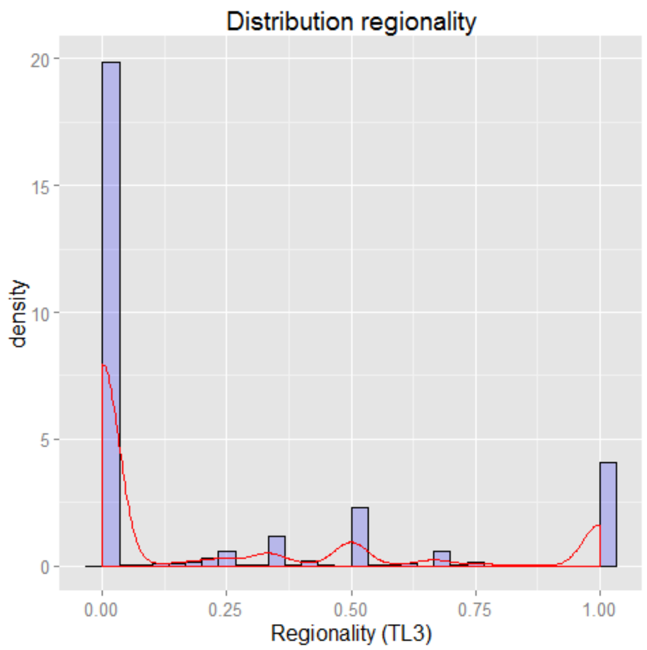
APPENDIX I: ADDITIONAL DESCRIPTIVE STATISTICS

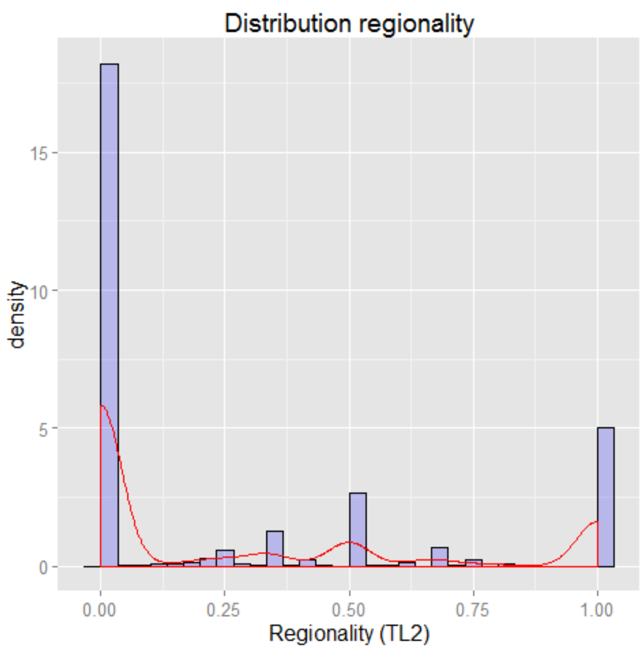
This appendix presents distribution plots of all variables (on log scale where appropriate), scatterplots of the measured relationships, and the regression models used.

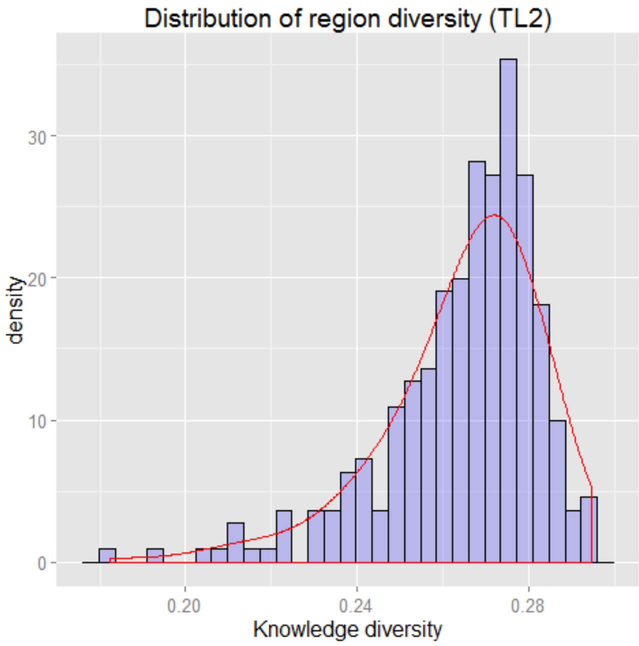
A: DISTRIBUTION OF VARIABLES

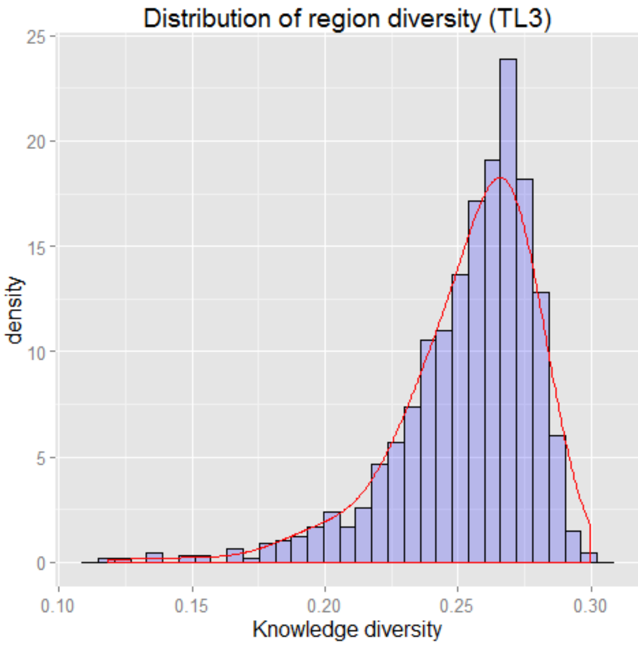


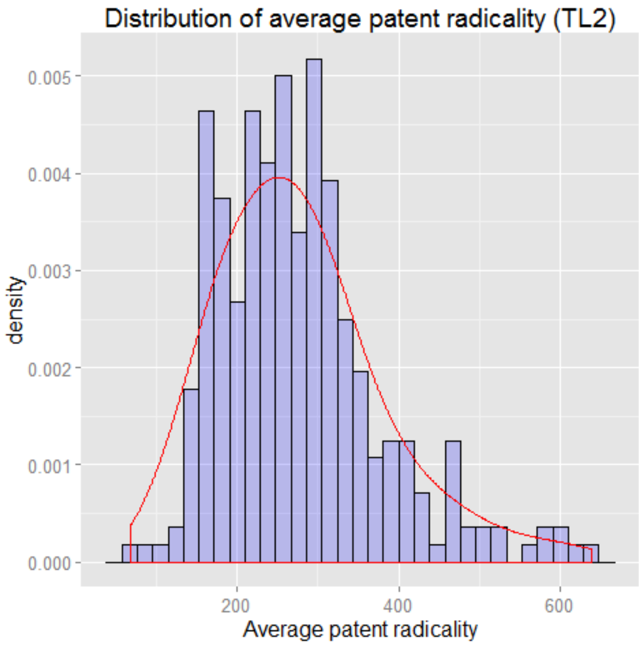


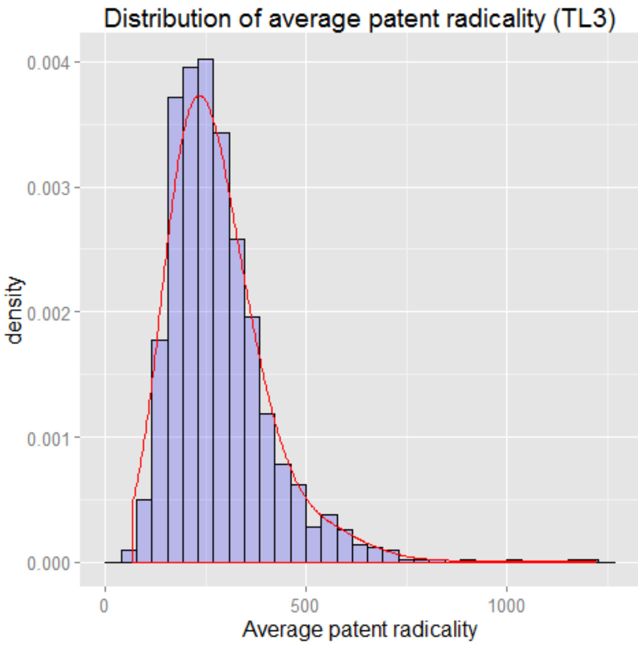


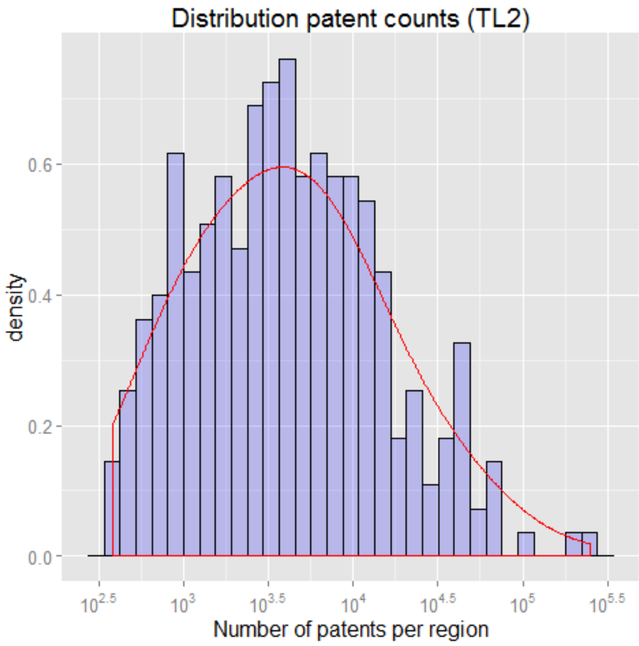


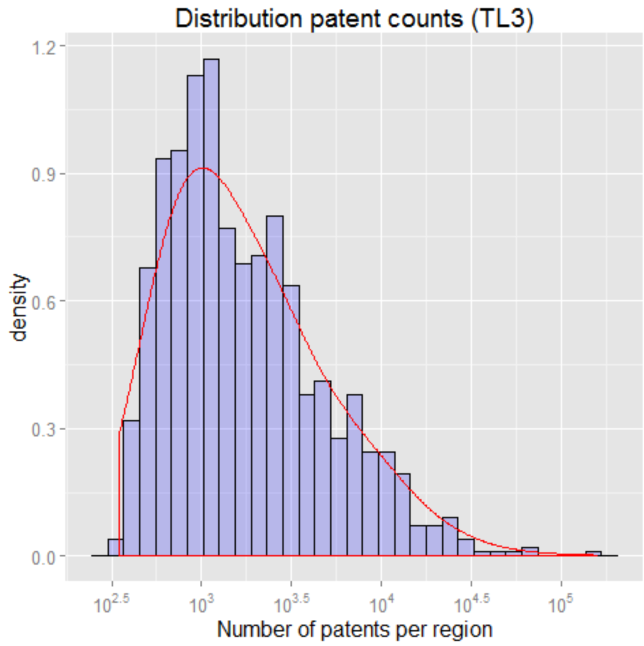

Distribution of region diversity (TL2)
density
0
10
20
30
0.20
0.24
0.28
Knowledge diversity

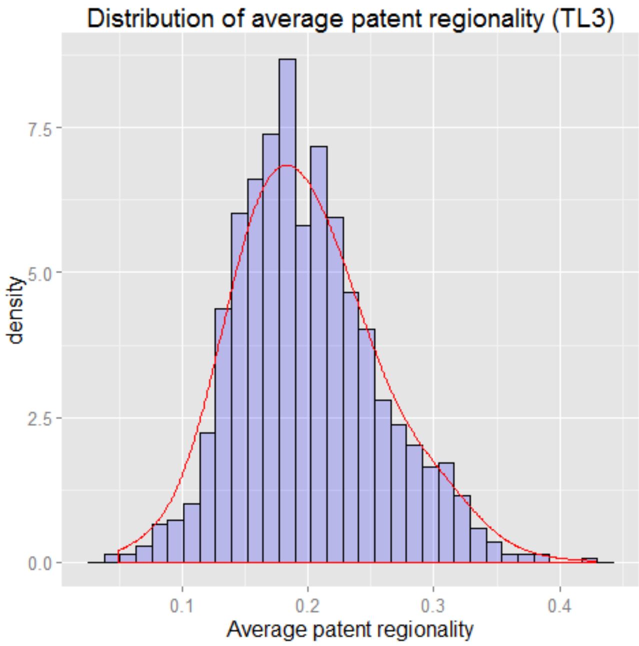

Distribution of region diversity (TL3)
density
0
5
10
15
20
25
0.10
0.15
0.20
0.25
0.30
Knowledge diversity

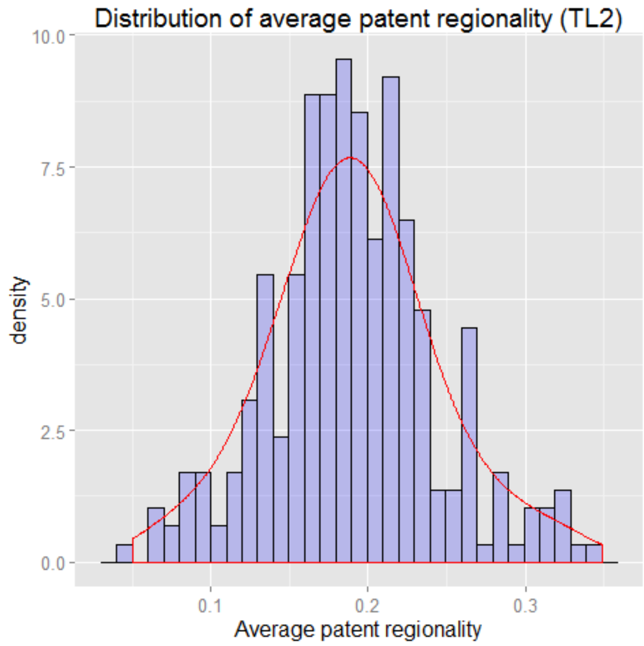

Distribution of average patent radicality (TL2)
density
0.000
0.001
0.002
0.003
0.004
0.005
200
400
600
Average patent radicality


Distribution of average patent radicality (TL3)
density
0.000
0.001
0.002
0.003
0.004
0
500
1000
Average patent radicality

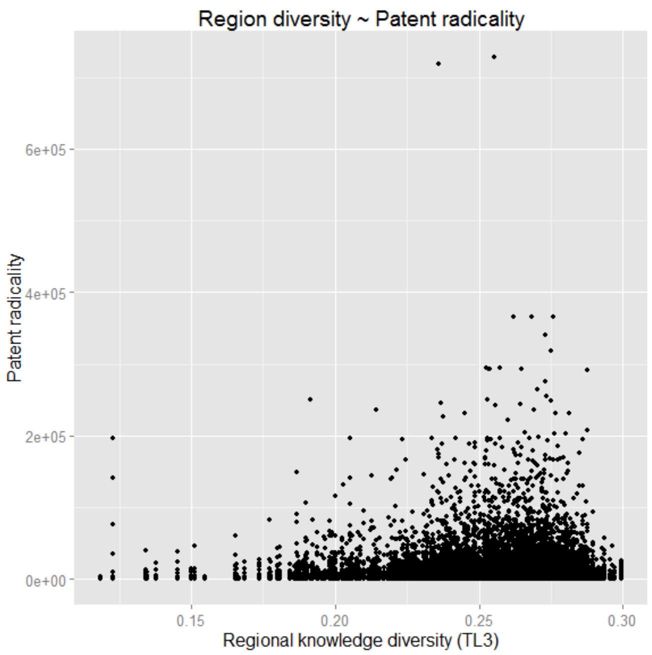

Distribution patent counts (TL2)
density
0.0
0.2
0.4
0.6
Number of patents per region

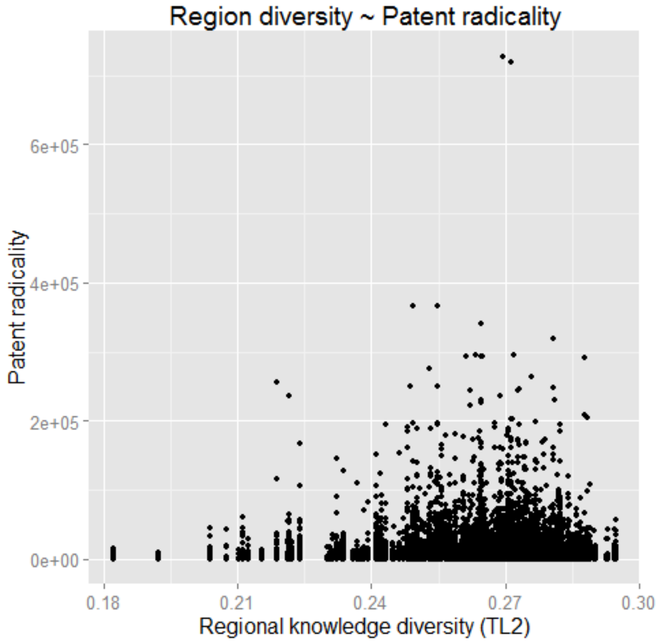

Distribution patent counts (TL3)
density
0.0
0.3
0.6
0.9
1.2
Number of patents per region

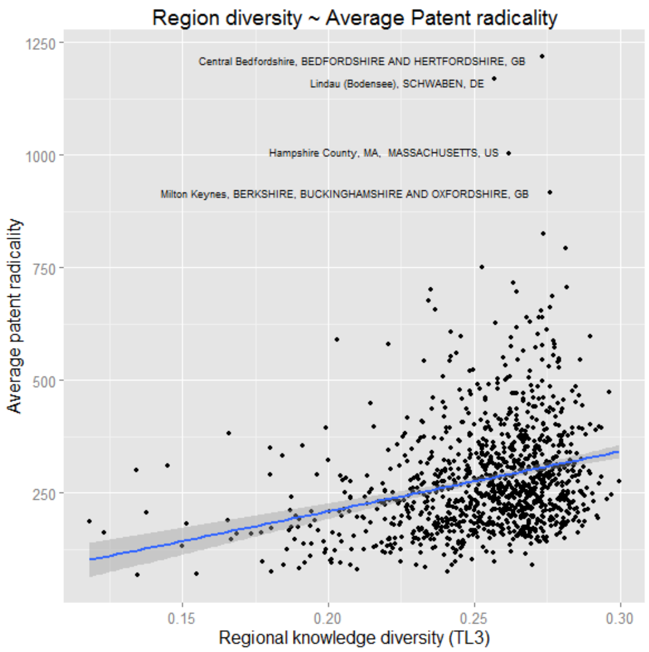

Distribution of average patent regionality (TL3)
density
0.0
2.5
5.0
7.5
0.1
0.2
0.3
0.4
Average patent regionality

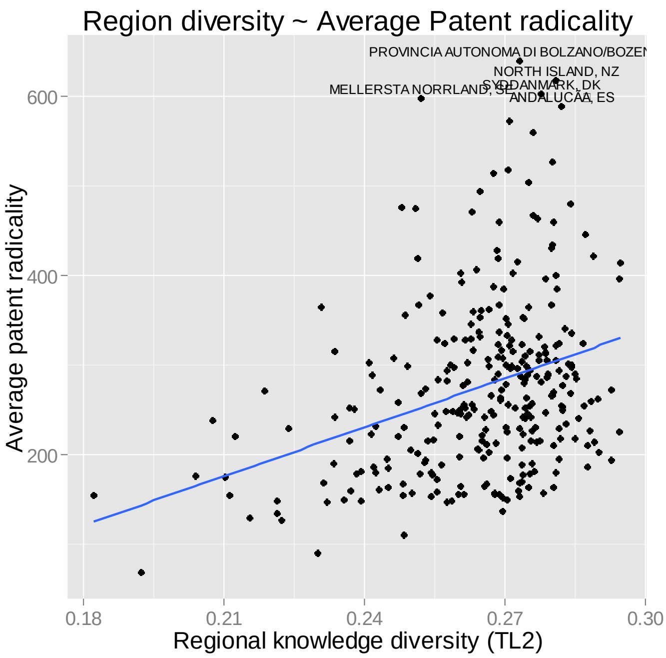

Distribution of average patent regionality (TL2)
density
0.0
2.5
5.0
7.5
10.0
0.1
0.2
0.3
Average patent regionality

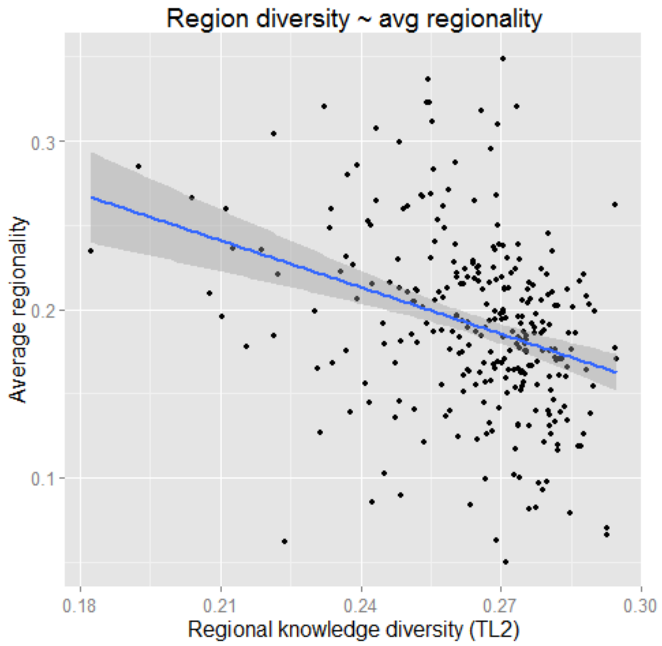
B: Scatter plots



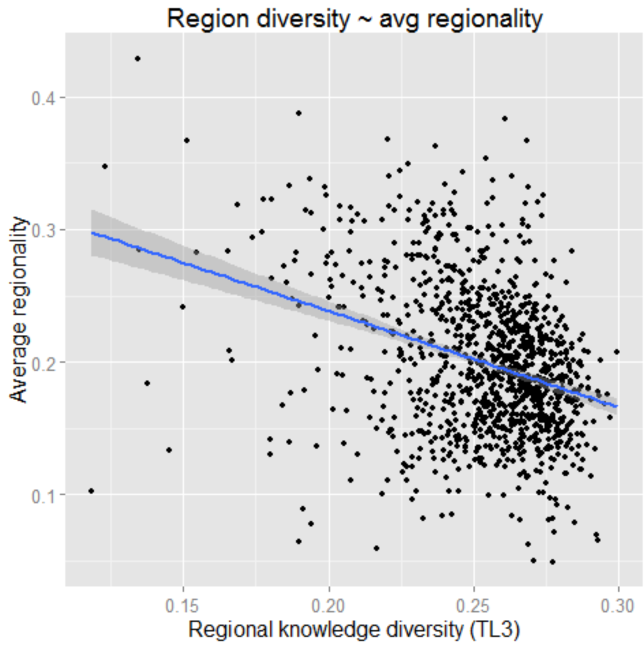








Region diversity ~ avg regionality
Average regionality
Regional knowledge diversity (TL2)
0.1
0.2
0.3
0.18
0.21
0.24
0.27
0.30



Region diversity ~ avg regionality
Average regionality
Regional knowledge diversity (TL3)
0.1
0.2
0.3
0.4
0.15
0.20
0.25
0.30

C: Regressions

Regression of radicality and region diversity on individual patent level

	Dependent variable:	
	radicality	
	(1)	(2)
reg_div_TL2	871.188*** (81.944)	
reg_div_TL3		763.848*** (48.406)
applicant_typeGOV NON-PROFIT	-5.463 (9.793)	-7.724 (9.794)
applicant_typeHOSPITAL	-153.530*** (42.064)	-155.215*** (42.063)
applicant_typeINDIVIDUAL	213.551*** (5.541)	210.912*** (5.545)
applicant_typeMULTI	27.627*** (2.436)	27.060*** (2.436)
applicant_typeUNIVERSITY	-43.610*** (8.943)	-43.860*** (8.943)
Constant	-20.852 (21.662)	15.333 (12.361)
Observations	3,639,324	3,639,324
R^2	0.0005	0.001
Adjusted R^2	0.0005	0.001
Residual Std. Error (df = 3639317)	2,186.636	2,186.595
F Statistic (df = 6; 3639317)	291.004***	313.678***
Note:		*p<0.1; **p<0.05; ***p<0.01

Regression of impact and radicality (quadratic)

	Dependent variable:
	fwd_cits5_xy
Radicality	-0.00001*** (0.00000)
Radicality^2	0.000*** (0.000)
Constant	0.218*** (0.001)
Observations	1,272,427
R^2	0.0002
Adjusted R^2	0.0001
Residual Std. Error	0.713 (df = 1272424)
F Statistic	95.947*** (df = 2; 1272424)
Note:	*p<0.1; **p<0.05; ***p<0.01

Regression of average regionality per region and region diversity

	Dependent variable:	
	avg_regionality	
	(1)	(2)
reg_div_TL3	-0.725*** (0.063)	
reg_div_TL2		-0.925*** (0.162)
Constant	0.384*** (0.016)	0.435*** (0.043)
Observations	1,101	294
R^2	0.108	0.101
Adjusted R^2	0.107	0.098
Residual Std. Error	0.053 (df = 1099)	0.050 (df = 292)
F Statistic	132.775*** (df = 1; 1099)	32.714*** (df = 1; 292)
Note:		*p<0.1; **p<0.05; ***p<0.01

APPENDIX II: INDEPENDENCE TEST COVARIATES

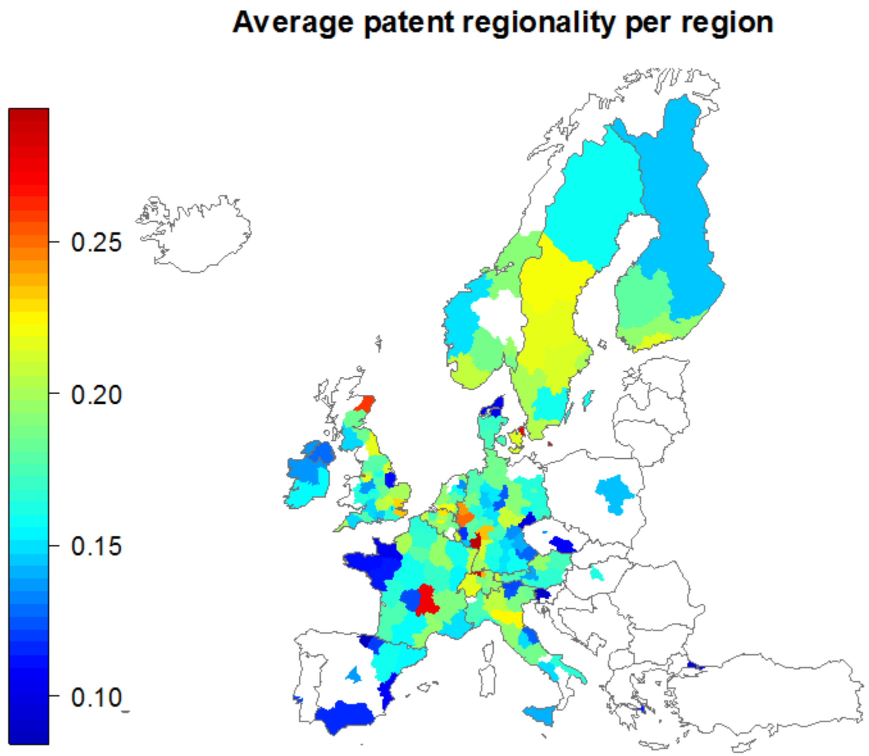
Covariates TL3 region level

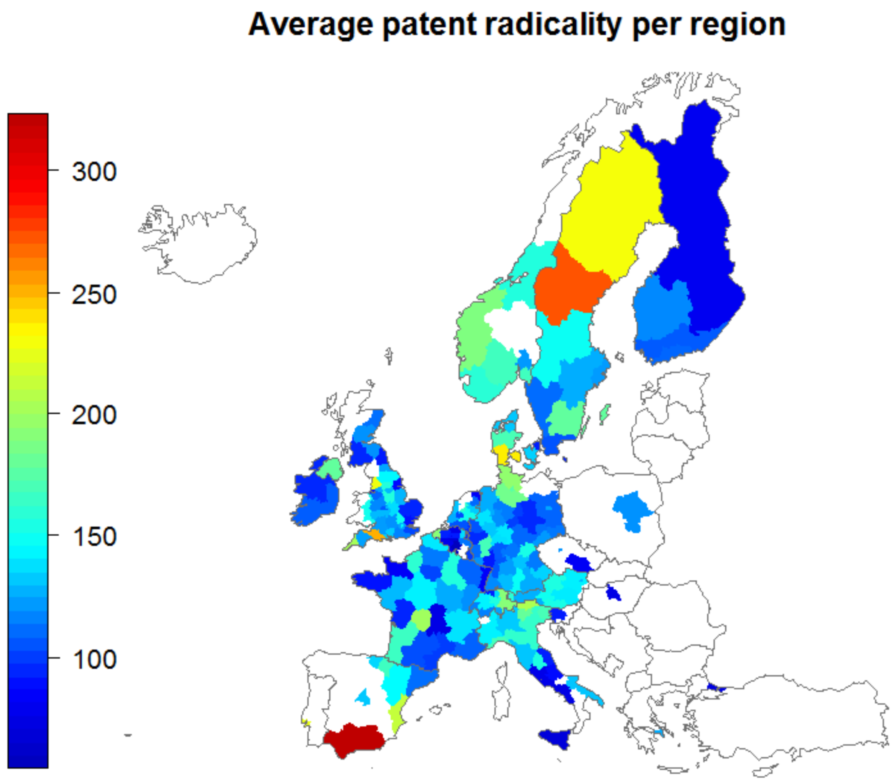
Pearson correlations								VIF
	npatents	reg_div	avg_radicality	COMPANY	GOVNONPROFIT	UNIVERSITY	INDIVIDUAL	
npatents	1	0.011	-0.194	0.047	-0.011	-0.029	-0.151	1.068372
reg_div	0.011	1	0.271	-0.239	0.136	0.041	0.342	1.220081
avg_radicality	-0.194	0.271	1	0.004	-0.057	-0.066	0.258	1.203824
COMPANY	0.047	-0.239	0.004	1	-0.407	-0.469	-0.571	2.843509
GOVNONPROFIT	-0.011	0.136	-0.057	-0.407	1	0.094	0.064	1.337621
UNIVERSITY	-0.029	0.041	-0.066	-0.469	0.094	1	0.035	1.502887
INDIVIDUAL	-0.151	0.342	0.258	-0.571	0.064	0.035	1	2.093183

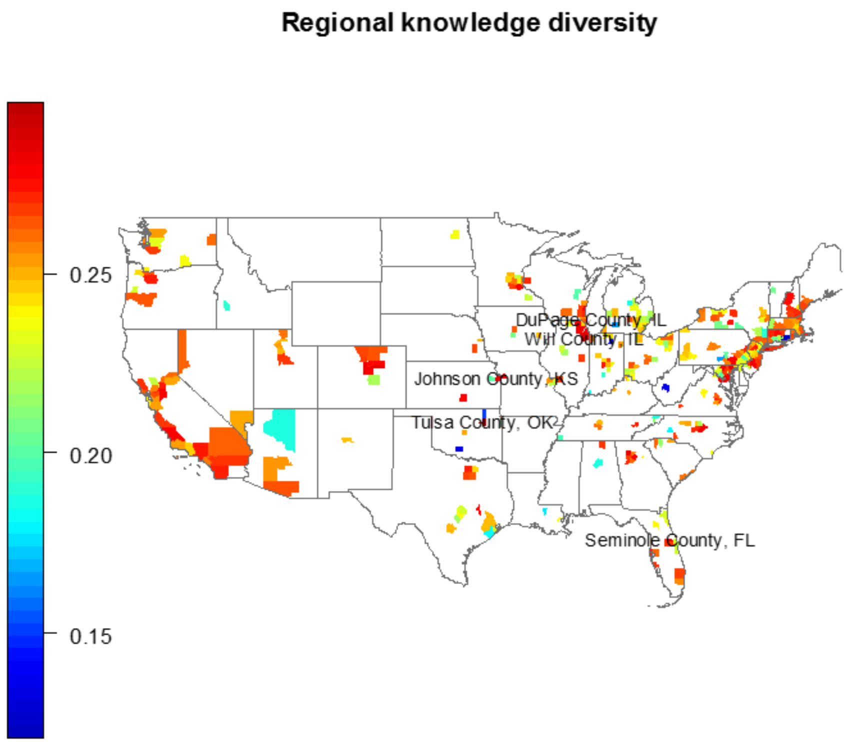
Covariates TL3 region level

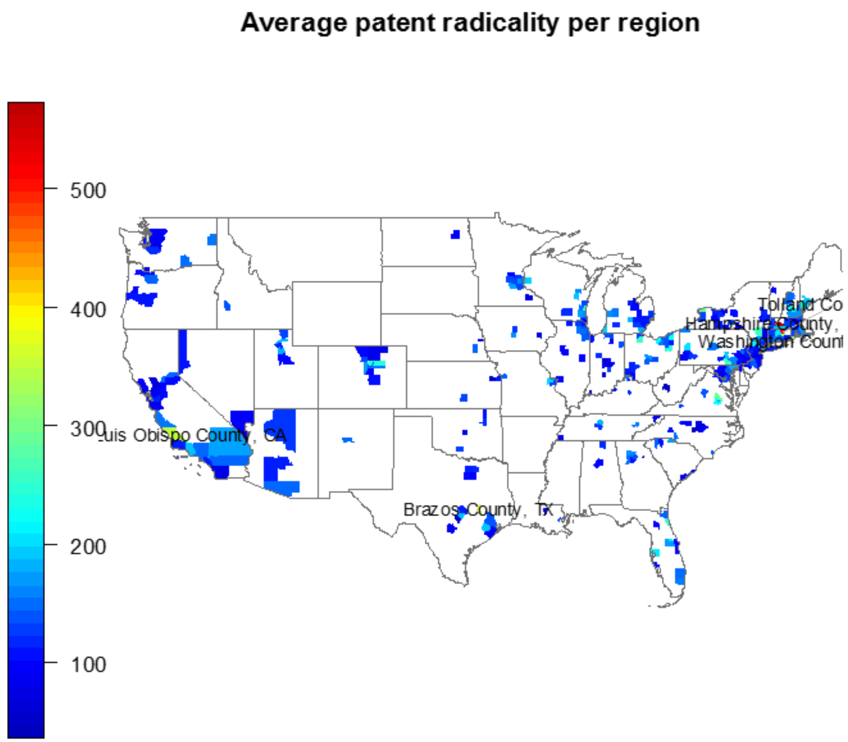
Pearson correlations											VIF
	npatents	reg_div	avg_radicality	GDP_PC	EDU_LF	POP_DEN	COMPANY	GOVNONPROFIT	UNIVERSITY	INDIVIDUAL	
npatents	1	-0.011	-0.226	0.179	0.188	0.120	0.212	-0.096	-0.092	-0.284	1.224833
reg_div	-0.011	1	0.343	-0.020	0.042	-0.033	-0.291	0.132	0.081	0.303	1.283972
avg_radicality	-0.226	0.343	1	-0.055	-0.224	-0.107	-0.067	-0.114	-0.157	0.373	1.452357
GDP_PC	0.179	-0.020	-0.055	1	0.288	0.352	0.089	-0.007	0.137	-0.055	1.331667
EDU_LF	0.188	0.042	-0.224	0.288	1	0.342	-0.281	0.245	0.141	-0.104	1.627234
POP_DEN	0.120	-0.033	-0.107	0.352	0.342	1	-0.115	0.146	0.025	0.168	1.393584
COMPANY	0.212	-0.291	-0.067	0.089	-0.281	-0.115	1	-0.481	-0.427	-0.594	3.522078
GOVNONPROFIT	-0.096	0.132	-0.114	-0.007	0.245	0.146	-0.481	1	0.206	0.081	1.488798
UNIVERSITY	-0.092	0.081	-0.157	0.137	0.141	0.025	-0.427	0.206	1	0.064	1.453921
INDIVIDUAL	-0.284	0.303	0.373	-0.055	-0.104	0.168	-0.594	0.081	0.064	1	2.612011

APPENDIX III: GEOGRAPHIC PROJECTIONS OF VARIABLES






Regional knowledge diversity
0.25
0.20
0.15
DuPage County, IL
Will County, IL
Johnson County, KS
Tulsa County, OK
Seminole County, FL


Average patent radicality per region
500
400
300
200
100
uis Obispo County, CA
Tolland Co
Hampshire County,
Washington Count
Brazos County, TX

X