EFFECT OF THE 2004 EU ACCESSION ON PATENT QUALITY IN THE CENTRAL EASTERN EUROPEAN REGION

By

Zsófia Zádor

Submitted to

Central European University

Department of Economics and Business

In partial fulfilment of the requirements for the degree of the Masters of Arts in Economic Policy in Global Markets

Supervisor: Professor György Bőgel

External supervisor: Balázs Lengyel

Budapest, Hungary

2019

Abstract

This thesis investigates how international collaborations with OECD nationals affect patent quality in the regions of the EU8 countries between 1990 and 2010. This was both before and after the accession to the European Union and the European Research Area (ERA) in 2004. To understand how international knowledge flows impact the previously more secluded regions of Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Slovakia and Slovenia, social network and econometric analyses are carried out based on the 2015 OECD REGPAT database. In the OLS and propensity score weighted difference-in-differences model radicalness is used as a quality measure. Radicalness measures how many different IPC technology classes the cited patents belong to. The findings show that regions of the EU8 have a 18.3 percentage point higher radicalness on average after treatment than the control group of Romania and Bulgaria. While joining the EU has a statistically significant positive effect on radicalness it cannot be regarded as international knowledge transfer, as is shown through the E-I index. Instead, the more inward-looking regions are, the better they are at patenting. International inventors are marginal in the collaboration network and thus are unable to create knowledge spillovers to other inventors they do not directly interact with. The seclusion of foreigners in the CEE region is supported by comparing the centralities of CEE and international inventors. These findings reveal that the benefits of the European Union lead inventors to better patenting and highlight the shortcomings of the ERA. In order to have successful knowledge flows it is not sufficient to only encourage the quantity of international collaborations. Foreign inventors also need to be embedded in the patenting network, so the innovation that is brought about by them can further spill over in the region and have a long-term effect.

Acknowledgements

First of all, I would like to express my gratitude to both of my supervisors. Professor Gyögy Bőgel, who helped me start my days thinking about my thesis with our morning consultations. And Balázs Lengyel, who guided me through this process, from learning the basics about networks to giving me comments on the final draft, but most importantly helped me see the value in all the work I have done. I thank both of their insightful comments and advice.

Secondly, I owe special thanks to the entire ANET-lab, with whom I spent coffee and lunch breaks discussing the state of my thesis and who were always happy to help me with coding, econometrics issues, literature recommendations and figure stylings.

Third, I would like to thank my wonderful peers at CEU, who were always there for me and for each other no matter how busy and overwhelmed we were. Special thanks to Ahmed Al-Shaibani, who has helped both my academic advancement and mental state these past two years.

Lastly, I would like to thank Gergő Tóth, without whom I would have never stumbled upon network science, the ANET-lab, my thesis topic or my econometric model. Thank you for inspiring me.

Table of contents

List of Figures

List of Tables

1. Introduction

As globalization is becoming more prevalent, the number and importance of research and innovation created in collaboration is also increasing. This trend is further supported by international bodies, that recognize the advantage of cooperation when it comes to knowledge production. Collaboration allows for knowledge transfer between different disciplines as well as on the same subject but with different background knowledge. Cooperation on projects help specialization that also allows researchers to gain a deeper understanding of their field. The output of such collaboration will thus be able to be more innovative, have a deeper impact and affect several fields at once.

Research and innovation are the engine of economic growth. With the help of research, technology can be improved that increases the value of future production, products and services. At the same time innovation is the basic function of entrepreneurship. Only with increased productivity, improved quality and reduced costs can the competitive advantage be improved. Not only does this help economic progress, growth and sustainability, but it also has an advantage for the population in general by increasing the standard of living.

In 1993 the European Union created an open market among its members, with free flow of goods, services, capital and people. The aim was to reduce transaction costs between member countries and with this increase the economic productivity of these countries. In 1998 when the Fifth Framework Programme was created with the aim of providing financial support for international collaboration, it became clear that knowledge flows are crucial for increased innovation. In 2000 during the Lisbon Agenda the European Research Area (ERA) was created with the explicit aim of furthering collaboration and mobility of knowledge workers.

Policy measures with this specific aim had to be implemented as physical and institutional distance are the main obstacles that innovators face when working in cooperation. Working together in person allows for the most effective knowledge exchange. Therefore, researcher mobility is still preferred to online communication. Taking this into account the ERA is working towards barrierless communication between researchers. Many research studies show the mobility of academics between research centers and across borders. What is still unclear is how the mobility of researchers increases knowledge flow and through this helps increase the quality of patents created in international collaborations.

One of the most important claims against using patents as a proxy for innovation is that not all of them are valuable. Therefore, measuring patent quality is necessary, but it has limitations, thus researchers still prefer to focus on the number of patents. The few studies that focus on patent quality include a study by Beaudry and Schiffauerova (2011), who find that the number of claims a patent has is higher for patents that were created in collaboration, and Lengyel and Leskó (2016), who show that the number of citations a patent receives is also higher in case of collaboration. Both claims and citations received are proxy measures for quality.

In this thesis I explore how a country's joining of the European Union affects inventors' quality of work. The EU8 countries are Central Eastern European countries that joined the EU in 2004 and thus have a clear-cut time when their borders opened and they began to experience the advantages of the single market and the European Research Area. The EU8 countries are the Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Slovakia and Slovenia. Because of this, these countries serve as good examples to measure the effect of increased international collaboration on patent quality, measured through radicalness. Radicalness counts the number of IPC classes, that the backward citations belong to. A patent is considered more radical if it relies on sources from many IPC classes (Squicciarini, Dernis, & Criscuolo, 2013). My research question is: how did the 2004 European Union enlargement effect the radicalness of patents produced in the EU8 countries?

I use patents as proxy measures for innovation and knowledge, which has been done by economists since before the mid 1950's (Griliches, 1990). Even if using patenting as a measure for innovative activity has flaws and may cause biases in measurement, they are still the most reliable and easily accessible form of quantitative research and development measures available. To carry out my quantitative analysis I use patent data from the 2015 OECD REGPAT database and OECD patent quality measures. I explore the research question first through social network analysis. Two networks have been created: a collaboration network between patent inventors and a collaboration network between regions. Econometric models are used to understand the quality change of patents in the EU8 regions. First a propensity score weighted difference-in-differences model is created, where the radicalness of patents produced is the dependent variable. The EU8 countries are the treatment group and they are compared to the control group consisting of Romania and Bulgaria. I chose these countries, because they are also part of the CEE region and were the next to join the EU in 2007. Thus, they are fairly similar to my treatment, but the 2004 EU accession has not affected them. Then the E-I index as a network measure is included in a regression to gain deeper understanding whether it is the international collaboration that has an effect on increased radicalness. The E – I index measures how in- or outward looking a region is based on the number of external and internal links (Krackhardt & Stern, 1988).

This thesis is a follow-up to the work by Arrieta, Pammoli and Petersen (2017), who found that an unintended consequence of the European Research Area was that the CEE region went through a brain drain effect. They estimate that had it not been for the ERA the region would expect larger cross-border collaboration and, while many have left the region, the return has been significantly lower. I build on their findings by looking at the quality of patents (that I identify with quality of knowledge) in the CEE region. My contribution is to further explore how the ERA affected the region. I found that joining the European Union had a statistically significant positive effect on patents created in the EU8 countries. But this is not a causal effect of foreign inventors bringing in new knowledge to the regions. Instead I find that more inward-looking regions, with higher number of internal link produce more radical innovations. The lack of spillover from international inventors may be due to the fact that in the network these inventors are marginal.

In the first chapter I will introduce the academic literature regarding pros and cons of using patents as indicators for technological change and economic growth, assess the ever-increasing importance of collaborations on patents and the success of ERA. In the second chapter I will introduce my research design, through information on patenting in the CEE countries compared to OECD countries, and an analysis on the CEE regions. I will then go on to the difference-in-differences method used to measure change in radicalness in the patents created in the EU8 countries and back up the relevance of factors with network indicators. In the last chapter I will draw inferences from my analysis regarding the ERA policy measures.

2. Literature review

In this thesis I am observing the quality change of patents in the Central Eastern European region with the accession to the European Union. While there has been work done on how patent quantity changes in the CEE region due to EU accession, very few papers can be found regarding the quality of patents due to the increased brain circulation. I also use patents as proxies for knowledge and innovation, as they have been used by economist for such purposes for decades. In order to access the liability of patents as proxies for knowledge I assess the pros and cons of carrying out qualitative research on patent datasets. One strength of such data is that collaboration of inventors can be observed through them. In our globalized world collaborations have been gaining significance and many advantages are regarded to such knowledge exchange between researchers. I turn to previous research to evaluate the kind of effect collaboration has on the quality of work, especially if it happens between inventors from different regions and countries. The CEE countries experience a sudden opening of borders and freedom of movement in 2004, that may lead to the increase of international collaboration, and brain circulation across borders. The European Research Area was established in 2000 to help the mobility of knowledge workers, but the success of it has to be evaluated.

2.1. Patents as proxy measures for innovation

The patent system was established to protect the intellectual property of inventors (Desrochers, 1998). Patents are documents that ensure the legal protection of inventions. These documents are granted by official State agencies; patent offices serve the purpose of ensuring that for a predefined amount of time only the inventor has the legal right to produce and use the new innovation. Innovations can include processes, physical equipments or machines alike. When patents are reviewed a large amount of information is gathered on them, regarding the inventor, the firm and what class the patent belongs to based on its technological properties.

Patents are useful proxy measures for technological change and knowledge growth as by definition they are the output of inventive and research activity. Patent offices grant or reject applications based on whether or not the filed patent is eligible and meets the given criteria. Because all patents need to pass the same criteria to be accepted, they can be regarded as objective measures, where standards are stable. Since patent offices grant a patent they gather extensive information regarding given innovation. Not only is there extensive information on a patent level, the systematic data collection is also available for decades. The data is thus readily available, cheap and a good source for quantitative analysis (O Lanjouw, Pakes, & Putnam, 1998). Patents are useful sources to measure collaboration and the analysis of networks and knowledge flows, because inventor collaboration is observable. Since technological knowledge exchange and knowledge flow cannot always take place by working in the same firm or same region, direct personal contact has to be observed in order to assume knowledge transfer. Patents are the direct output of collaboration and thus show us a more reliable flow of ideas and information on an inventor level.

Using systematically gathered patent information is easy but has weaknesses and disadvantages when drawing conclusions from quantitative analysis. First and most importantly, not all innovative activity results in the filing of a patent, but even the ones that do may differ greatly in how original, radical and economically valuable they are (Acs, Anselin, & Varga, 2002; Desrochers, 1998; Griliches, 1990). Many research points to the law of large numbers as an assurance that when measuring patents different quality inventions will be included in the analysis. Desrochers (1998) also points out that bias regarding patenting may be a result of exogenous factors, such as the size or ownership of the firm, the concentration of innovative centers and the industry in which the innovation has been created. Desrochers (1998) raises further issues that are a result of incorrect data gathering and identification of classes that a patent belongs to. Further disadvantages arise when we test how well patents measure what economists expect to measure with the use of patenting data. Ács and coauthors (2002) came to the conclusion that even though the extent of knowledge flow on a regional level is exaggerated patents are still relatively reliable measures of exchanging ideas. Hall, Jaffe and Trajtenberg (2001) on the other hand raise issues with measuring the quality of patents with citations and citation related measures (such as originality and radicalness). As they find that there is a changing trend in the amount of citations over time and between patents belonging to various fields.

Patents are thus fairly reliable but indirect measures of innovation, but when using them economists have to be aware of the limitations regarding data collection and identification as well as biases of measurement. Patents can be used as indicators of proxies for innovative activity but are not suitable to measure all kinds of research and development and economic growth. Therefore, it is appropriate to use information on patents to analyze knowledge flows, knowledge gain and information transfer in this thesis, as no overarching deductions regarding economic growth are made.

2.2. The effect of collaboration on patents

Creating patents in collaboration has been increasingly important that can be seen through the growing number of collaborations throughout the years. Working together on a project helps the division of labor, deeper understanding and elaboration in one's own field of interest, a new level of understanding of ideas coming from different fields. Interchanging knowledge and methods and creating a new technology in this manner ensures that the patent will have a more radical impact on various fields at once. Collaboration also helps ease the financial cost of training of inventors in areas they are not familiar with as with collaboration they can either learn from their co-inventors or allow them to take over tasks that they are more familiar with.

The European Union recognizes the need to further collaboration and mobility of inventors and have been implementing programs, which are aimed at promoting economic activity, innovation and job creation. The European Parliament takes a crucial role in developing the European policy for research and technological development (RTD) through three main platforms; Framework Programmes, international collaboration and the European Institute of Innovation and Technology ("Policy for research and technological development | Fact Sheets on the European Union | European Parliament," n.d.). The main objective of the RTD is to improve competitiveness of the EU by encouraging technological development and research capacities with the help of free knowledge flows that can be achieved through scientific collaboration.

The Framework Programmes (FPs) were created to provide monetary funding to carry out the European policies for science and technology. Since the First FP in 1983, these programmes became primary advocates of international scientific collaboration. The Fifth Framework Programme (Commission of the European Communities, 1997) introduced in 1998 was the first ambitious program aimed at closing the gap between EU and competitors regarding technological development. One of the key points of this program is unlocking the resources of the living world; this is approached partly from researcher mobility, collaboration and ensuring that less developed rural and coastal areas are integrated within the research community. The objective of the program is "interoperability and interworking of diverse infrastructures" (Commission of the European Communities, 1997, p. 22) through technological and infrastructural development. Shortly after in 2000 the European Research Area was formed during the Lisbon European Council meeting, aimed at establishing the free movement of knowledge workers, devices and processes within the EU (Commission of the European Communities, 2000). It carried on the goal of the Fifth Framework Programme, but was more directly focused at collaboration and the exchange of ideas and knowledge, by bringing both researchers and research institution into closer contact.

Scientific collaboration among international researchers has grown significantly on the global sphere throughout the years as globalization has helped open up national borders. OECD countries are in the process of globalizing their technologies (Archibugi & Michie, 1995), but this is more prevalent for large technologically advanced countries (Guellec & van Pottelsberghe de la Potterie, 2001). On the global level, Wagner, Park and Leydesdorff (2015) find that within two decades the number of internationally coauthored scientific papers have increased by more than double, but while connections are denser, they do not cluster. Even though there is a general trend to increase international collaboration and the EU is investing in furthering knowledge transfers, European Union countries still struggle with the mobility of knowledge workers (Chessa et al., 2013; Picci, 2010).

Inventor mobility and international collaboration are frequently studied subjects as they carry knowledge flows within themselves (Bathelt, Malmberg, & Maskell, 2004; Boschma, 2005; Hoekman, Frenken, & van Oort, 2009). Knowledge transfer can be observed through the quality of patents created in collaboration, or through the quality of patents produced by foreign firms in the home country. International collaboration leads to more innovative research, this can be seen through the higher number of claims received by internationally created patents in Canada (Beaudry & Schiffauerova, 2011) and the higher number of citations received by international patents in the CEE region (Lengyel & Leskó, 2016). New inventors entering a research center or firm result in combined knowledge, furthermore the mobility of knowledge workers also brings dynamics to the existing status quo knowledge production. Not only do mobile inventors contribute with new ideas to the patent they are creating and the inventor group, but often a spillover effect is observed, that helps knowledge flow to other inventors and firms in the region (Breschi & Lissoni, 2001; Guan & Chen, 2012; Jaffe, Trajtenberg, & Henderson, 1993; Varga & Schalk, 2004).

While international knowledge exchange is vital for radical inventions, collaboration through space (national and administrative borders) remains a challenge. Morescalchi et al. (2013) affirmed that distance and physical borders have a negative effect on inventor collaborations. When looking at OECD countries they find that distance has a decreasing effect on the number of collaborations only until mid-1990's but starts to gain strength later. This is surprising especially in the light of technological advancements that were expected the reduce the significance of geographical distance and bring people together. Therefore, if cross-country collaborations are more challenging to carry out, inventor mobility is of vital importance that ensures personal interactions, which was found to be more effective regarding knowledge transfer (Anselin, Varga, & Acs, 1997; Balland & Rigby, 2017; Jaffe et al., 1993; Storper & Venables, 2004).

2.3. Researcher mobility in the European Research Area

Since joining the European Union allowed researchers of the CEE region to move freely across borders, as well as Western inventors to enter the countries of the 2004 accession, I expect that there will be higher number of collaborations. International collaboration increases the quality of patents produced; mobility is also expected to have a positive effect in technological advancement in the region. EU accession on the other hand, does not only allow freedom of movement for the people, but also goods, investment, reduces transaction costs across borders and improves the overall assessment of the countries.

In case of Central Eastern European countries, that started opening their borders towards the west only after the dissolution of the Soviet Union, international academic collaborations are of special importance (Fitjar & Huber, 2015; Grillitsch & Nilsson, 2015; Marzucchi, Antonioli,

& Montresor, 2015; Montobbio & Sterzi, 2013; Penrose, 1973; Varga & Sebestyén, 2017). Inventor mobility has both its advantages and disadvantages in the CEE region, on the one hand, it allows for international researchers to enter these countries more freely, and domestic inventors to travel and gain experience and knowledge abroad that can be later invested in regional innovation. On the other hand, it also allows for an easier way to leave for technologically more developed regions. The Central Eastern European region has experienced a loss of human capital and brain drain, since joining the EU and opening of markets. But at the same time it also experiences brain circulation that furthers the development of good quality innovations (Petersen & Puliga, 2017). At the same time, the number of patents created in the region is lower than if the countries had not joined the EU (Doria Arrieta, Pammolli, & Petersen, 2017).

The creation of the European Research Area has not brought about the expected changes regarding knowledge worker mobility within the European Union. There is still a need to harmonize administrative and institutional requirements, as both geographic and institutional distance have a negative effect on collaborative patenting (Hoekman et al., 2009). Furthermore cross-border collaboration has not intensified since the establishment of the ERA (Chessa et al., 2013). Instead increased inventor mobility is likely to lead to an increase between-country inequality (Grossmann & Stadelmann, 2011).

No evidence was found regarding any significant positive effect on the intensity of collaboration between regions, institutions or individuals, since the establishment of ERA (Chessa et al., 2013). While there is an upward trend in the number of collaborations that take place, this however increases with the same speed as do collaborations in other OECD countries. Chessa et al. (2013) find an increase starting from the 1990's that seems so stall in 2003, three years after the establishment of the ERA.

3. Data

3.1. Data overview

The dataset used in this thesis is from the publicly available 2015 OECD REGPAT database, that contains information about patents on a regional level from 1970 until 2010. There are 2,737,669 unique patents in the data, and 4,981,173 unique inventors, but as inventors collaborate on patents and one inventor often has many patents, the total data consists of 6,957,520 rows. I merged it with a 2013 OECD database on measuring patent quality and regional data from EUROSTAT on NUTS3 level (Nomenclature of Territorial Units for Statistics). This allowed me not only to look at the dispersion of patents but also how the quality varies in time and space with regards to demographic and regional differences. I narrowed my data down to patents that were created either in the 1) EU8 countries 2) Romania and Bulgaria or the 3) OECD countries. For general data description I concentrated on the 1990 – 2010 period, but further narrowed down the data for my econometric analysis to 1999 – 2010.

I used the 20-year period for description to gain an overall understanding of patenting trends regarding quality and number of inventions, but since my benchmark is at EU accession in 2004 it was reasonable to look at equal periods both before and after the threshold. After filtering, the dataset contains 978,280 unique patents and 1,458,862 unique inventors, and a total of 2,216,973 rows. My research question was focused on the Central Eastern European region; on the EU8 countries that entered the European Union in 2004. The EU8 countries will be compared throughout the thesis to the two other groups, the OECD countries (except for those, that belong to the EU8 group) and the 2007 EU entrants: Romania and Bulgaria. Out of the almost a million patents 10,192 were created by at least one of the 20,132 EU8 inventors. The Romania Bulgaria control group consists of 1383 inventors and 755 patents.

The dataset I created from the above-mentioned sources allowed me to examine patents, inventors, regions and countries at the same time with the appropriate aggregation measures.

The OECD REGPAT is an inventor level database, which contains information on which patent each inventor contributed to, as well as which country and region they are from. Inventor location is given on the NUTS3 level. The OECD quality dataset is on the patent level and has 21 different quality measures and the year of filing. The EUROSTAT data is on the NUTS3 level, which contains the most detailed regional information. I gathered data on the area of the regions and population density. I created two more datasets from this information, one that is patent level and the other that is region level. For the patent level dataset, I collected information from my merged data regarding the number of inventors, whether or not the patent was created in collaboration and whether the patent was created with an OECD inventor. On the regional dataset I congregated information on the number of patents and inventors in each region each year, and average quality measures.

From the merged dataset two networks were constructed, a co-inventor network of those patents, where at least one of the inventors is from the EU8 countries and a regional network, where at least one of the inventors produces at least one more patent after international collaboration. These relationships were defined on the 1990 – 2010 period and were weighted on the number of patents two inventors or regions work together.

3.2. Quality measure

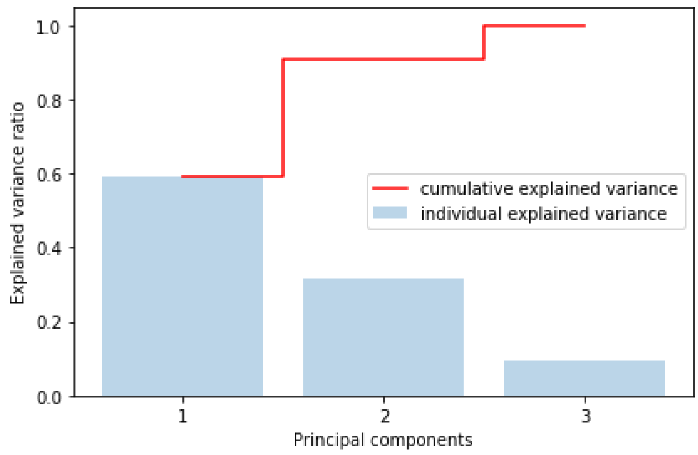
Those opposing the use of patents as measures of innovation claim that not all patents have the same quality and often companies safeguard one groundbreaking idea, by creating patents for all the small inventions that total up to the breakthrough patent. To account for the variations in patent quality it is important not only to look at the number of patents produced, but the quality itself. A few quality measures have been used by economists to measure the importance of innovation, most often the number of citations a patent receives, but the number of claims also shows up in some analyses. The main issue with these measures is that they vary across technological fields and firms. Realizing the need to measure technological and economic value of patents without strong bias, OECD came up with a wide set of quality measurements, that can be used for comparison through space and time (Squicciarini et al., 2013).

Squicciarini et al (2013) develop indicators for patent measures that can be used as proxies for innovation. Their work is of high importance since they use information from patent documentation and compute quality measures that consider shocks in patenting. The authors create normalized indexes, that take into account the quality of other patents created in the same year and in the same technological field. This ensures that scientific fields that use more claims, cite more or use in general more sources do not have a positive bias compared to fields where such steps have a smaller importance. The homogeneity of these measures thus allows for reliable cross-country analysis.

For my analysis it was important to find a quality measure that is reliable in the less documented Central Eastern European patents as well. So, when analyzing the measures, I first excluded the ones with many missing values in my area of observation. Secondly, I excluded all measures that could not have been properly determined as they rely on data after the patent has been documented, for example, various forward citation measures, that measure the number of citations a patent got in the upcoming years of documentation, or patent renewal.

Radicalness, patent scope and number of fields a patent belongs to were chosen and a Principal Component Analysis carried out. Figure 1 shows, that after the orthogonal transformation the radicalness variable can explain 60% of the variance. As can be seen through the cumulative explained variance patent scope still explains a more significant variance of 30%, while the number of fields patents belong to seems not to carry a high variance. At the same time the sign of correlation between the three quality measures are not the same. That is why I do not create a principal component from the transformation of the three measures. The strength of radicalness in the PCA still provides support for using this measure in the analysis.

Figure 1 - PCA on patent quality measures



Source: Author's own visualization using Python

Radicalness ranges from zero to one, the higher the measure the more scientific fields the patent uses knowledge and inspiration from. According to the official OECD working paper: "Radicalness of a patent is measured as a time invariant count of the number of IPC technology classes in which the patents cited by the given patent are, but in which the patent itself is not classified." (Squicciarini et al., 2013, p. 53). The authors further normalize this measure by the number of IPC classes that can be found in the backward citations, thus it can be calculated as:

$$Radicalness_p = \sum_{j}^{np} \frac{CT_j}{n_p} \text{ ; } IPC_{pj} \neq IPC_p$$

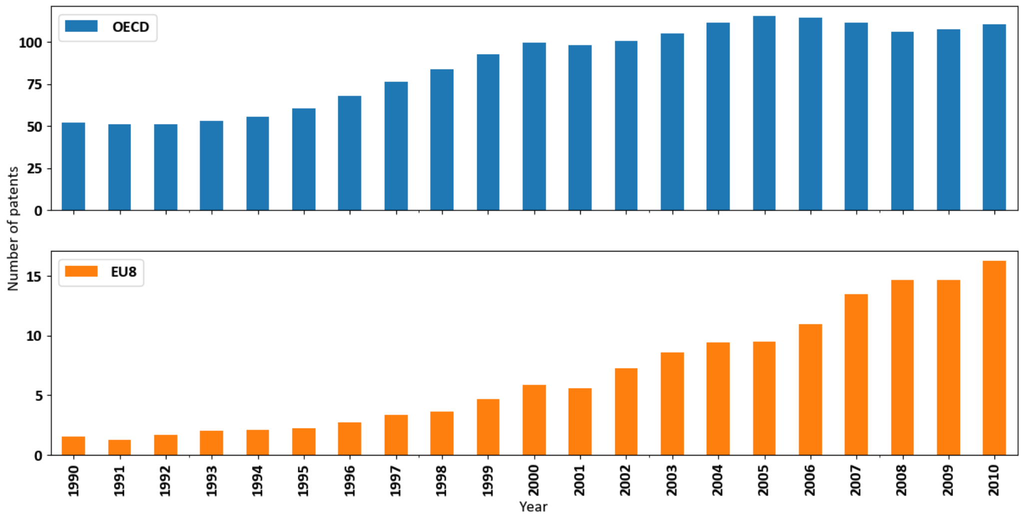
where CT is the count of 4-digit IPC classes in the backward citation, n is the total number of IPC classes in the backward citations in the field the patent of observation belongs to, p is the patent we are looking at, and j is the patent cited by the patent of observation (Squicciarini et al., 2013).

The number of different IPC classes patents rely on is generally increasing over time and across technological fields. With the denominator accounting for the total number of IPC classes the technological field variance can be strained, but the general upward trend will still remain. The normalization with n_p is also important as it gives an overall understanding of how well the patent is performing compared to other innovations of its field.

3.3. Country level description

I explored the original, unfiltered dataset to gain a basic understanding of the trends and the relevance of the expectations. I looked at patent quantity and quality, grouped by years and countries. Figure 2 shows us the number of patents over a million citizens invented each year from 1990 till 2010 in the OECD and in the CEE countries. When looking at the scale, it can be observed that OECD countries produce a substantially larger number of patents than CEE countries, even when I scale with population. In 2010 the OECD countries produced 140,000, while CEE countries created only 1200 patents, when I scaled the measures, the difference between the two still greater than 100 patents. There are differences between the two histograms trend wise as well, while the patent invention in the OECD countries experience a sharp increase from 1994 – 1995, they reached a more discrete development in the beginning of the 2000's. The CEE countries experienced a more exponential increase in the observed 20-year period. Interestingly in 2004 there was a stall in the number of patents produced in the region, while I would have expected a sudden increase. This stall might be due to the fact that new entrants had to undergo institutional adjustments, that slowed down the pace and the administrative tasks of patenting. Another stall can be observed from 2008 to 2009, which might be an effect of the crisis, although interestingly there is a larger increase in the number of patents the year after.

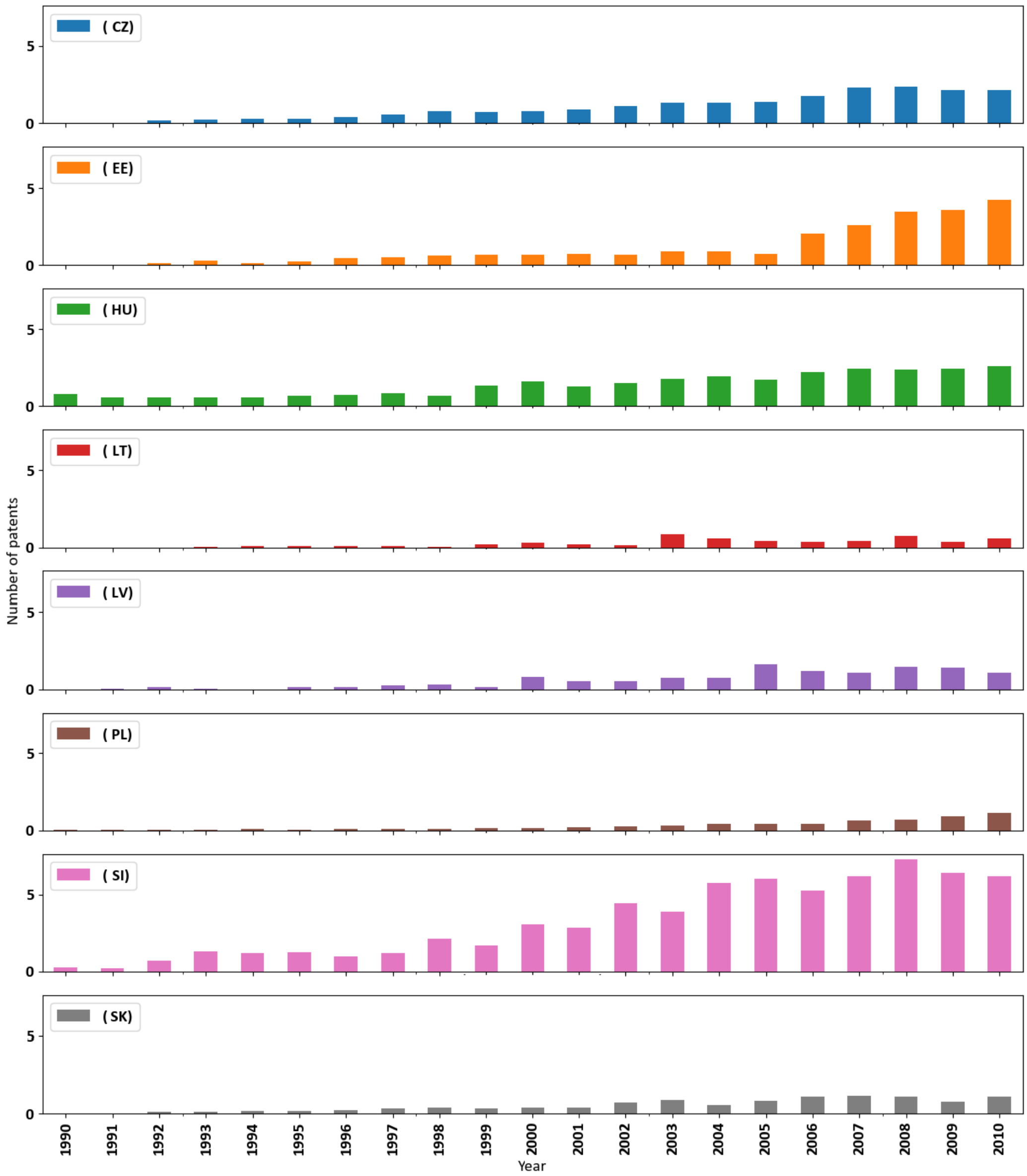
Figure 2 -Number of patents over a million citizens per year



Source: Author's own visualization using Python

Figure 3 shows the number of patents per 100,000 people in each of the EU8 countries. In this detailed figure I converted the scale to patent per 100,000 people so that not only trends can be detected but more meaningful information regarding innovative activity can be also perceived. For this purpose, I adjusted the scales to the same measure, maximum 6 patent per 100,000 people. Figure 3 shows us that there are observable differences in innovative activity in the CEE countries. By 2010 Poland produced 400 patents a year, the highest number of patents in 2010 produced by a CEE country. Despite of the high number of patents, when per capita measures are explored Poland is among the worst performing countries in the region, together with Lithuania, Latvia and Slovakia. On the other end of the scale, there is Slovenia, who outperforms all other nations in per capita patenting, and does well regarding overall number of patents as well. Hungary and the Czech Republic produce both high number of patents (up to 200 a year in each), and do well innovation wise when we scale with population. Estonia is also outstanding in per capita patent production, even though the overall number of patents produced is insignificant. So, for many CEE countries there is a clear distinction between per capita patenting and number of patents produced. Slovenia is outstanding in per capita patenting, followed by Estonia, while Latvia, Lithuania, Slovakia and Poland are the least performing, regarding per capita patenting.

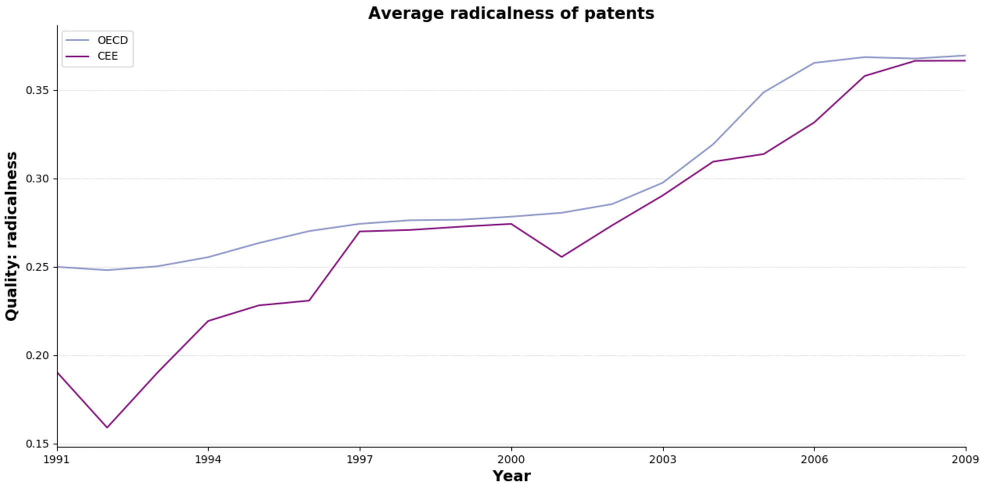
Figure 3 - Patents per 100,000 people in the EU8 countries



Source: Author's own visualization using Python

I checked the quality of patents in two different ways. First, I looked at the average radicalness each year when the patent was created by OECD inventors and compared them to patents created by at least one CEE inventor. Figure 4 shows an overall increasing trend regarding radicalness of patents. This means that with time inventions are getting more innovative and are gaining knowledge from a larger pool of scientific fields. The blue line represents average radicalness of OECD patents, that start at an average of 0.25 and increase by 0.1, so a 10-percentage point increase takes place over the observed period. For CEE patents the quality development is more significant, radicalness increases by more than 15 percentage points, and by 2009 they outpace OECD patents.

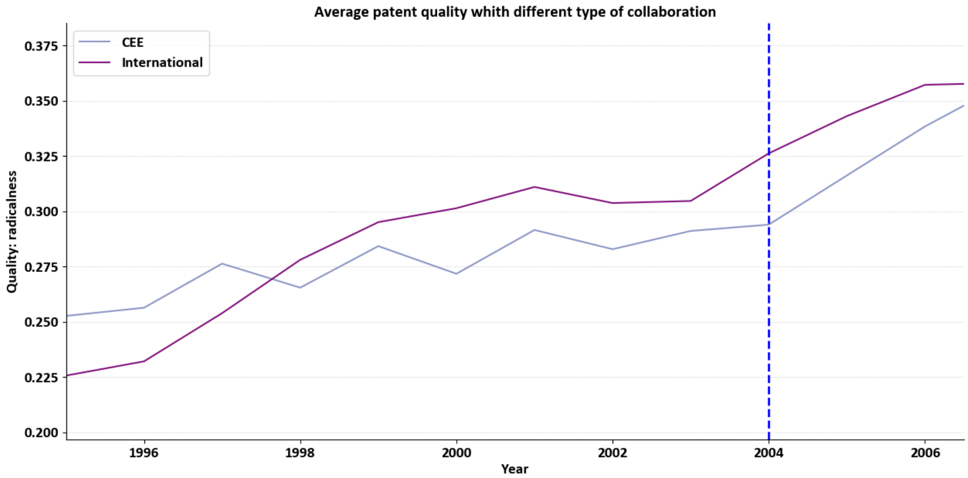
Figure 4 - Average radicalness of patents per year



Source: Author’s own visualization using Python

Secondly, I explored how patent radicalness changes in time, when they were created only by inventors from the CEE region or in international collaboration. Figure 5 shows that the average quality difference is not significant between the two groups, but when observing the purple line - the patents created in international collaboration – a more than 13 percentage point increase can be seen, compared to the 9-percentage point increase in CEE patents. Not only does this show that collaborations improve radicalness more rapidly than inward looking collaborations, it also suggests that with the opening of the market better quality collaborations take place, with inventors, with more embedded knowledge. In both groups, radicalness stagnates in 2007, a slow down in development that could be related to the 2008 economic crisis.

Figure 5 - Average patent quality with different type of collaboration



Source: Author's own visualization using Python

3.4. Region level description

I aggregated the data to regional level because the purpose of investigation is to find spatial patterns and regions also ensure heterogeneity. While I could have chosen a country level comparison, doing so would not take into account either size and population of the countries, or regional disparities. Comparing patenting abilities of Estonia to Romania, for example would be a questionable method. At the same time regional differences within a country can further help in understanding the socioeconomic mechanisms that take place. Hungary for example has a low number of regional centers, that perform significantly better than the urban areas. But regional comparison can also help in addressing division, such as Germany's east-west, or Italy's north-south divide.

Furthermore, a regional level analysis is crucial for the methodology used in this thesis. Both the difference-in-differences and the External – Internal index can only be carried out properly, when data is grouped by regions. NUTS level regions are classified in a way to ensure homogeneity of population size, but also give information on degree of urbanization, whether the region is mostly urban, intermediate or rural. Innovation is explored spatially and density of research center and firms is an important factor in knowledge spillover. For the difference-in-differences model I need observations in two periods, both before and after the 2004 accession, but most inventors do not patent in the two periods. Instead of losing all information on patents, whose inventors only patented once, I aggregated the information on a regional level for each period. This way I got the average radicalness of patents in the two periods for each region. The E-I index is a measure that looks at the number of links within a group, and links going to outside groups. When exploring spatial patterns, regions are the most obvious measures for groups, so that links between inventors from the same region and from different regions can be taken into account.

I chose the most detailed NUTS3 level regional classification to carry out my econometric and network analysis. EUROSTAT defines 3 levels of regional classifications, NUTS level 1, 2 and 3, from largest to smallest (Brandmüller, Önnerfors, Reinecke, & Statistical Office of the European Communities, 2018). The different NUTS levels are defined based on population size, therefore smaller countries are not divided on level one or two. Since the subjects of my analysis are the CEE countries, that involves Estonia, Latvia and Lithuania, countries that are not yet divided on the NUTS2 level, it makes sense to use NUTS3. NUTS3 level regions need to have a population between 150 – 800 thousand, while NUTS level 2 regions have a population of 800 thousand – 3 million. While NUTS3 levels may provide disparities in major countries, with a population size of millions (such as London, that is divided to 20 NUTS3 regions), for my area of study it is a more suitable size of measure.

4. Method

4.1. Network analysis

I created a bipartite patent network; whose constituents are inventors (in different countries and regions) and patents that are created by them. Cooperation arises between different inventors, so they can share knowledge and create something that could not have been possible alone, or only through immense resources. It can be thought of as knowledge or information flow between people. The distribution of patents is broad, with a long right tail, as only a few inventors form a large number of patents, while many inventors create patents only a few times. This is why, the rich get richer rule can be examined. The more inventors cite the patent, the wider its scope, and the more original it is the more inventors will base new ideas on them. Thus, forward citations (citations that a patent receives) are permanently giving feedback on the usefulness and quality of patents and knowledge of its inventors. The bipartite network has directed links from inventors to patents. If inventors work on a patent together once, they are more likely to work on another one again in the future. Collaboration between inventors is likely, but some patents are also formed by only one inventor.

From the bipartite network a collaboration network with only inventors can also be created, that maps out the flow of knowledge. As it is a social network it can be described by social network characteristics, that are distinctly different from other kinds of networks. In the collaboration network many triangles are likely to be present (Balland, Belso-Martínez, & Morrison, 2016; Giuliani, 2013; Juhász & Lengyel, 2018). If inventor A works together with inventor B, and inventor B works together with inventor C, A will be more likely to work with C, since there is a direct link to reach and collaborate with her. There are likely to be hubs of acknowledged inventors, who are more often asked to join projects (Boschma & Fornahl, 2011; Wal & Boschma, 2011). Geospatial characteristics are also likely to have an effect on the network, as inventors far away have less chance to know each other and cooperate thus several more separated groups will be present with links to other groups (Bathelt et al., 2004; Glückler, 2007). Collaboration networks are common forms of social network analysis, often used to measure knowledge flows in scientific articles, patents or through citations.

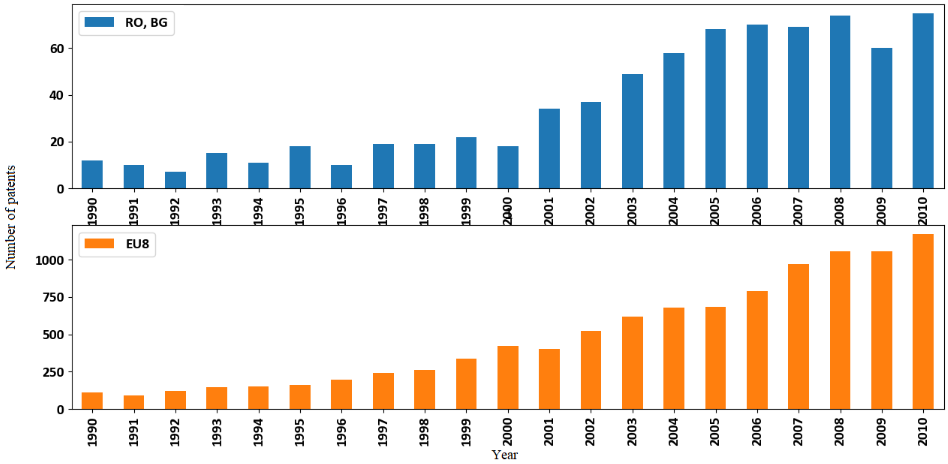
Two networks were created. I filtered the inventor – patent two-mode as described in the Data overview chapter and transposed the data to turn it into a one-mode collaboration network. The projection leads to loss of information, for example the collaboration network will no longer show us on which patent the inventors have worked together; it will only keep information on how many times such collaboration has happened. Second, I created a regional network, in which I gain information on inventors that are able to pass on the foreign information they gained during international collaboration. In this one-mode network I explore where foreign knowledge comes from and the likelihood of spill-over effect due to regional inventors continuing patenting.

4.2. Econometric analysis

Two econometric models – a propensity score weighted difference-in-differences and an OLS – were carried out. The goal was to understand if accession to the EU really improves patenting quality in the affected regions and whether this positive change can be due solely to the advantages of joining the EU, or it can be directly connected to the increased number of international collaborations made possible with the help of the European Research Area.

9000 more patents were created in the EU8 countries compared to Romania and Bulgaria. Among the EU8 countries Hungary, Poland and the Czech Republic have the highest number of patents, compared to the other 5 countries that were part of the 2004 enlargement. When observing the scales on Figure 6, it can be seen that while the number of patents created in the two groups is very different, the upward trend is similar. It is smoother for the EU8 countries, while Romania and Bulgaria experience a greater jump in the number of patents produced from 2000 to 2001.

Figure 6 - Number of patents per year in Romania and Bulgaria and the EU8 countries



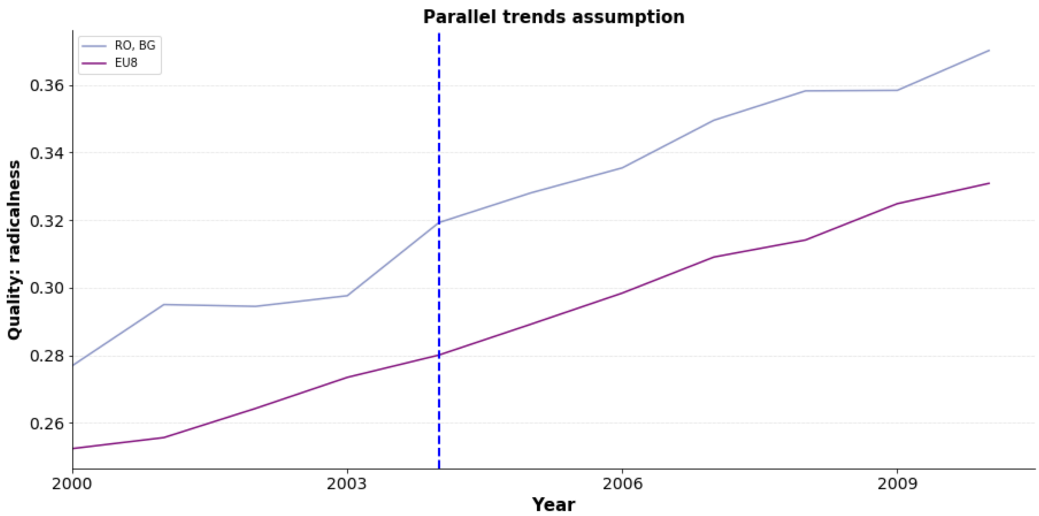
Source: Author's own visualization using Python

The size difference between the control and treatment groups was substantial and thus causes outliers to affect the control group more than it does the EU8 countries. When I aggregated my data to the region level, I found 57 regions in Romania and Bulgaria that created patents in the given time period and 145 regions in the EU8 countries. As described in the Region level description chapter, this aggregation was necessary both due to data constraints and to ensure heterogeneity. One limitation is that it is preferable for the control group to have a larger number of samples, since during matching treatment regions are compared to controls. Because of this not all regions can be included in the model later.

4.2.1. Difference-in-differences

I am examining how entering the EU affected radicalness of patents, and since the EU8 countries all entered at the same time in 2004, a clear threshold can be determined in that year. Romania and Bulgaria are used as a control group as they followed accession to EU shortly after the EU8; in 2007. The three-year difference is not large, suggesting a similarity in level of development between countries of the 2004 and 2007 enlargement. Since I have both control and treatment with a set of policy changes that only one of the groups experienced, a difference-in-differences model can be used. The parallel trends assumption, that ensures that both treatment and control groups experienced similar environments and policies seems to hold true based on Figure 7. Surprisingly the control group has slightly higher quality measures throughout the observed period, compared to the EU8 countries. The trend before 2004 is not completely parallel, since the number of observations for the control groups is a lot smaller, extreme observations and outliers affect the average measures more, resulting in a less smooth trend. The quality of patents is increasing at a higher level until 2001 but then stagnates for two years, then from 2004 the average radicalness is more smoothed out. At the same time the EU8 countries experience a straight monotone increase from 2001, with a slight drop in 2008. The average difference between 1999 and 2010 between the two groups is present, but is not large, on average the control group has around 0.02 higher radicalness measure than the treatment.

Figure 7 - Parallel trends assumption regarding radicalness measure of patents



Source: Author’s own visualization using Python

While the treatment and control groups are similar in a lot of sense even before 2004, they were not completely alike, as different governments made different measures to catch up with the west. We have to assume that there is also a reason why Romania and Bulgaria were not allowed to join the 2004 enlargement. To account for these differences, I aggregated my data to regional level with average radicalness measure. I introduced a matching between comparable units, in this case NUTS3 regions. Matching takes observable variables (X) and conditions on them, as if they were random. This way only those nontreated observations will remain for the difference-in-differences to take place, that are similar enough to the treated observations, and have strictly positive propensity scores. I used a kernel estimator as matching algorithm, that estimates for the probability density function of treatment observations and according to these distributions, weights control observations closer or further from the treatment distribution. As covariates I used population density, region size, number of inventors in the region during the period, number of patents in the given period and average number of inventors, who work on a patent.

Apart from the parallel trends assumption, a kernel-based propensity-score matching model was developed to ensure causal interpretation. I used this model to exclude the possibility, that regions within the EU8 countries differ from Romanian and Bulgarian regions in a way, that has a direct effect on the average radicalness of patents. This model sets assumptions for identification and is preferable to a simple diff – in – diff since matching prior to the model helps with functional form misspecification and makes more robust estimates.

There are limitations to this model, first the number of observations in the control group are smaller than those in the treatment, and thus not all treated regions can be included in the model, only those where a feasible match is found from the control group. There is also a stronger bias if the X, that includes the covariates are not suitable for matching, if there are unobservable characteristics that affect quality. For example, researchers may be happier to move to regions with pleasant environment, that may not be measurable by population, size, scientific innovation or wealth. Such characteristics may be weather, geographical location, proximity to sea or mountains.

4.2.2. OLS

Secondly, I created an Ordinary Least Square regression (OLS), with average radicalness of the region as a dependent variable, and used the covariates from the difference-in-differences model together with the EI network indicator to predict it. The OLS was carried out on the same dataset as the difference-in-differences method and provided further insight to what drives patenting radicalness in the Central Easter European region.

I calculated a network indicator called E-I, that measures how outward looking a certain group is and with the indicator I carried out two OLS estimations. The E-I index is based on the number of external and internal links as defined by Krackhardt and Stern (1988):

$$E - I\ index = \frac{External\ links - Internal\ links}{External\ links + Internal\ links}$$

The E – I index is defined on a range of -1 and 1. The closer the index is to 1, the more external links the group has, and the more outward looking it is. In this analysis I define my groups as regions, and internal links are edges with inventors, who patent in the same region, while external links are edges with inventors from any regions but my observed.

5. Results

Through my network analysis I saw that collaboration between inventors is highly fragmented and no one giant component can be found. The number of inventors who keep on patenting after interacting with a foreigner is also very low, hindering knowledge spillover. Despite the marginality of foreign inventors, the result from the difference-in-differences method shows an increase in radicalness of the EU8 regions compared to a control group of the 2007 EU entrants. When looking deeper into this result I found that better quality patents are not the result of regions collaborating with other regions, or showing outward looking relations.

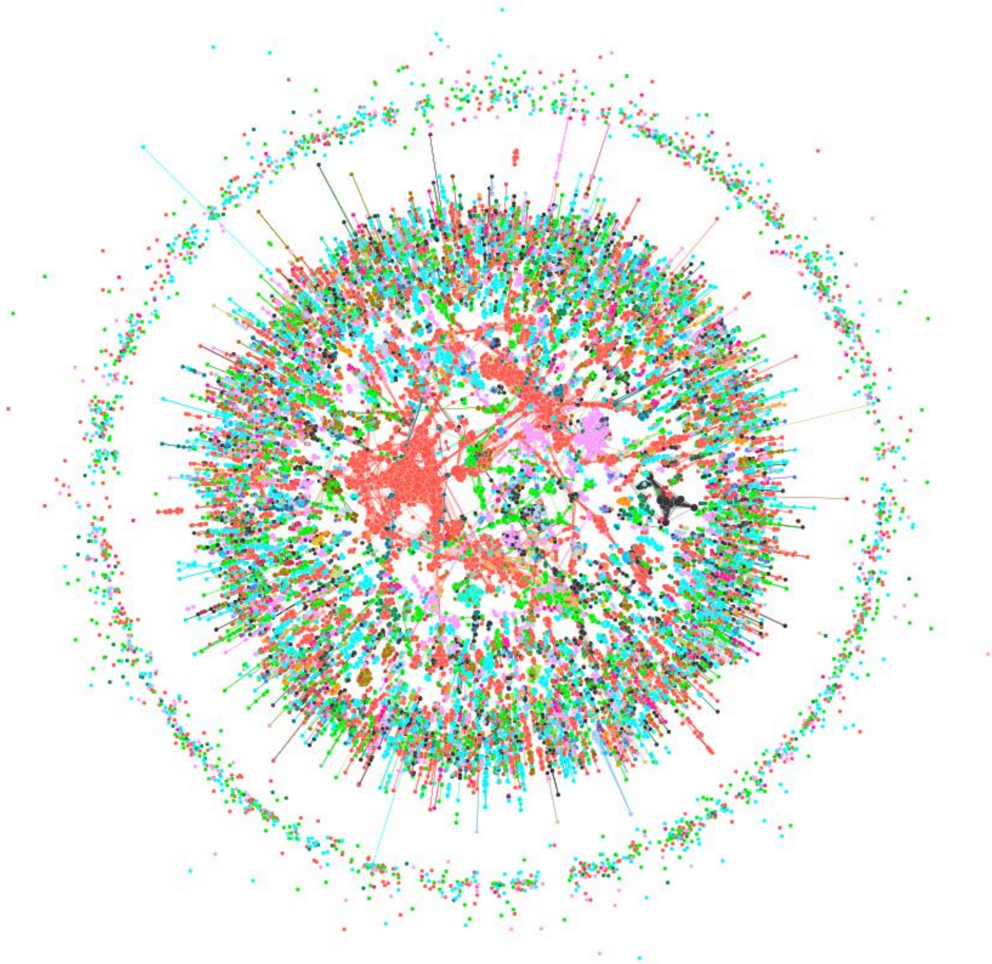
5.1. Network analysis

The inventor collaboration and regional collaboration networks both point out that international collaboration is rare, not well embedded in the network and thus does not allow for further spill over of the foreign knowledge transferred by them.

5.1.1. Inventor collaboration

The collaboration network of inventors working on patents in the Central Eastern European region was created from a bipartite graph of inventors and patents. The network consists of 27,706 nodes and 73,670 edges. The nodes are inventors who worked on a CEE patent any time between 1990 and 2010, and there is an edge between inventors if they have worked together on a patent. The average degree in this network is 5.32, meaning that the average inventor has worked with 5 other inventors in the 20-year period. On Figure 8 the whole collaboration network is presented and the graph is drawn so, that highly connected components are in the center and less connected ones are on the periphery. Figure 8 shows a thick periphery of unconnected nodes: these are inventors who are lone patentors and do not collaborate at all in the examined period. The center of the graph shows a highly connected component, consisting almost exclusively of red nodes. 27% of the total is red; this color represents Hungarian inventors. The second largest national group is the turquois Polish inventors, who seem to be a lot less connected to each other compared to the Hungarian ones, but still represent 17% of the total nodes. Similarly, Czech inventors are also frequent but more fragmented on the graph, with a 13% light green occurrence. On the other hand pink nodes, that represent Slovenian inventors, are highly clustered again, even though only 8% of the CEE inventors are from the country. The full network consists of 6,736 components, again showing how fragmented the network really is.

Figure 8 - Collaboration network



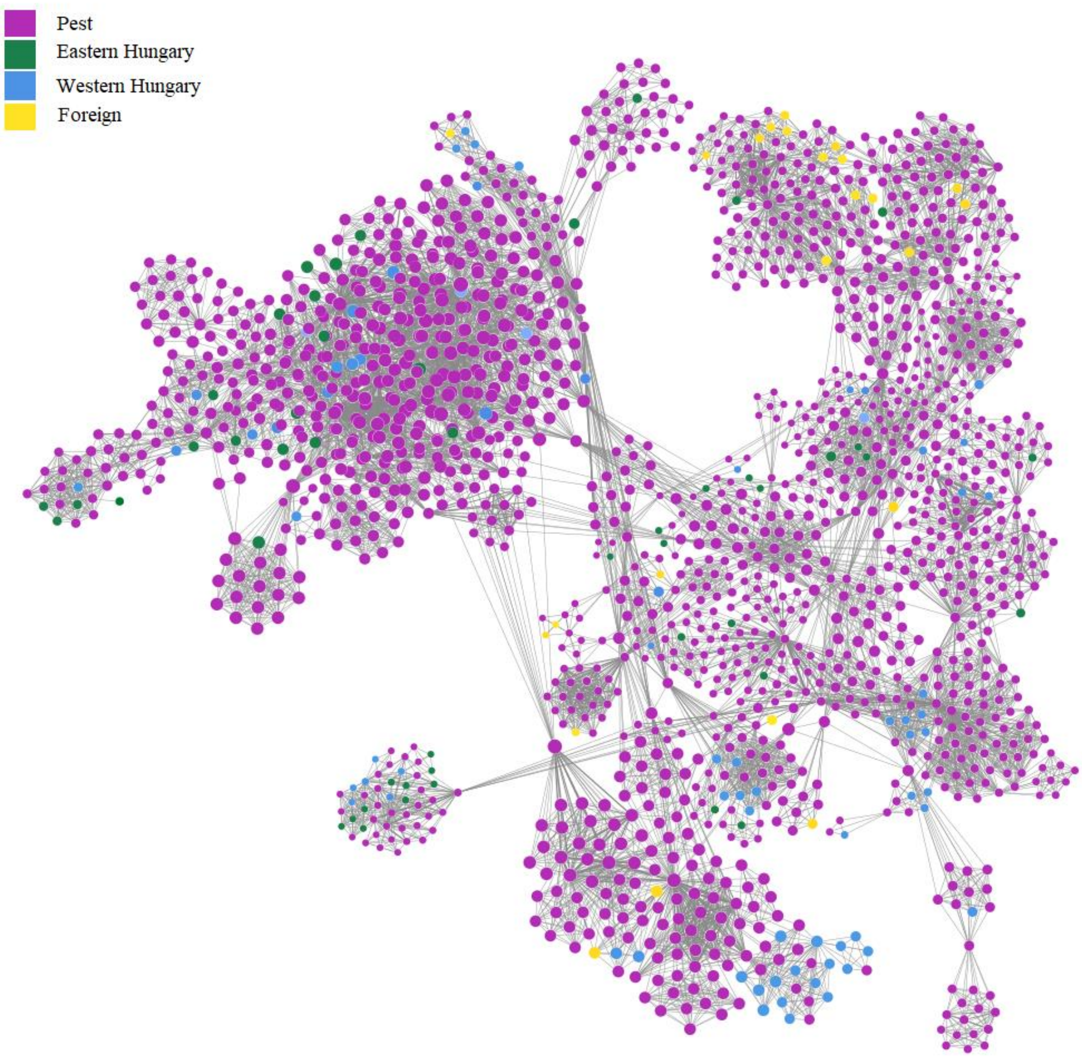
Source: Author's own visualization using Gephi

There is a substantial difference regarding the largest connected component and the number of lone patentors on the periphery of the network. This is due to the discrepancy in the type of institutions inventors patent. Small, national firms and research institutions are less likely to patent in large collaborations, while multinational firms motivate their employees to create patents this way. Therefore, large connected components in Hungary and Slovenia are likely to be present due to the high number of foreign owned firms that entered the market. These firms were already familiar with the importance of innovation and patenting, so when they invested, they created incentives for their employees to patent.

In order to get a deeper understanding of the network structures, I looked at the two largest connected components of the network. Components are sub-graphs, in which any two nodes are connected to each other by a path. The largest component of the network consists of 1,652 nodes and 11,797 edges, with an average degree of 14.3. Compared to the average degree of the full network, it can be seen in Figure 9 that the largest connected component is a lot denser, with a high number of inventor collaboration. This subgraph consists of connections only between Hungarians and OECD nationals, no collaboration happens across CEE countries. The graph could be divided to two clusters, which are only connected by a few brokers. On the left side of the graph I find an extreme density, and high number of edges, while the right side of the graph is a larger component, with fewer connections within. The yellow nodes represent foreign collaborators, and the purple, green and blue colors represent the area of Hungary the inventors come from. Most inventors work in the Pest region, but Eastern and Western Hungarian inventors are present too.

The lack of embeddedness and number of foreign inventors is crucial when knowledge is taken into account. From the 1,652 inventors only 24 are foreign. 13 of these OECD inventors are from France; the rest are from the USA, Canada, Italy, Israel and Sweden. What Figure 9 shows is that foreign inventors are marginal in the network and are often not highly connected to Hungarian inventors. The left side component contains only one foreign patentor, but not even this inventor is in the highly dense area of the network, but instead is only connected to a low number of inventors. The right-hand-side component has a higher number of foreign inventors and as can be seen on the top of the graph, they are also closer to each other. This might be a cluster of the network that can feel the effect of the foreign knowledge.

Figure 9 - Largest connected component



Source: Author's own visualization using Gephi

Centrality measures are used to identify important nodes within a network, and different measures show the the various tasks a node can take up in a network. I looked at betweenness, eigenvector and degree centrality. Betweenness centrality calculates how often a node is a bridge between two communities, eigenvector centrality measures the importance of a node, by assuming that connections to important nodes are more meaningful, and degree centrality calculates importance of a node based on the number of degrees it has. I calculated average centrality measures for CEE and foreign (OECD) nodes and present the results in Table 1, which 1 shows that all average centrality measures are higher for CEE inventors than OECD inventors. OECD inventors do not create bridges between other inventors, do not connect to nodes with high degree centrality and have on average 4 less connections than CEE inventors.

Table 1 - Centralities of the largest component

	Degree	**Betweenness**	**Eigenvector**
OECD	9.7	0	0
CEE	14.36	0.0047	0.009

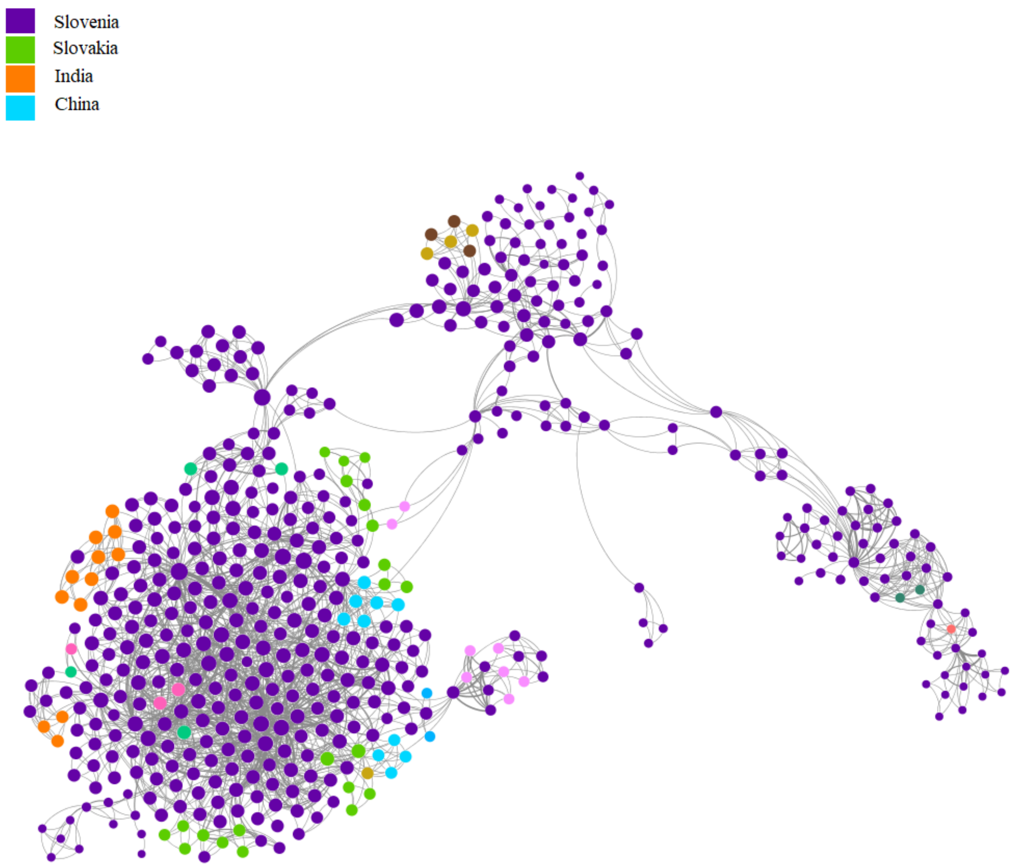
Source: Author's own calculations using Python

The second largest connected component is presented in Figure 10 and consists of 505 nodes, 2042 edges and have an average degree of 8. In contrast with the largest component, this one shows collaboration between the EU8 countries as well as with OECD inventors. 86% of nodes are Slovenian inventors, represented by purple, but 4% of Slovakian inventors are also present on the graph. This component has more foreign collaboration with inventors from 10 different OECD countries, the highest number of these inventors is from India, China, Austria and Switzerland. All together 45 inventors are from OECD countries, 11.2% of the inventors in the graph. Again, similarly to the largest component presented on Figure 9, most connected, central nodes are in general from one CEE country, in this case Slovenia. Slovakian inventors are on the periphery as well as OECD inventors. Compared to the largest compontent, Figure 10 shows a few foreign inventors who are more connected and in the center of collaboration.

Slovenian inventors are likely to be in the centre, since they are permanent employees in large multinational firms, who have more knowledge on what their co-workers' expertise is.

Working together and having information on your co-workers's specialization reduces transaction costs of patenting. This is a comparative advantage of national inventors, compared to foreigners, who have to invest in finding the right person to collaborate with.

Figure 10 - Second largest connect component



Source: Author's own visualization using Gephi

When I examined the average centralities of CEE and OECD inventors in the second largest component, I found that CEE nodes have a higher importance regarding all three centrality measures. The results are summarized in Table 2, and show that while OECD countries have lower betweenness and eigenvector centralities, and collaborate on average with 1 less inventor, the difference between the two groups is not as outstanding as in case of the largest component. The smaller difference might be partly due to the fact that more international inventors are present in the network, and partly due to the higher number of them.

Table 2 - Centralities of the second largest component

	Degree	**Betweenness**	**Eigenvector**
OECD	6.51	0	0.01
CEE	8.25	0.00089	0.02

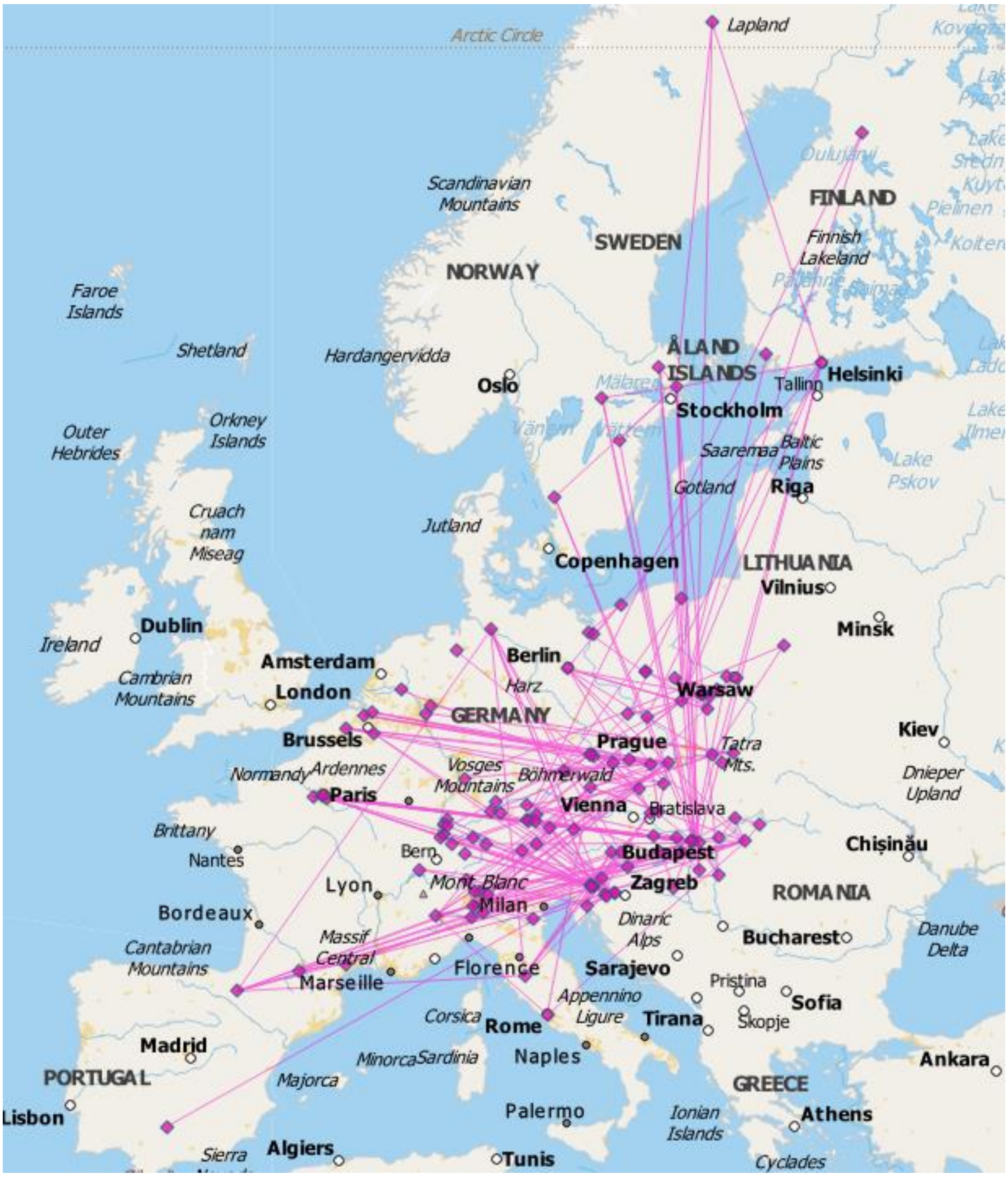
Source: Author's own calculations using Python

5.1.2. Regional collaboration

My second network plots regional collaboration of those inventors who collaborate with an OECD inventor once and keep patenting after at least once. The network consists of 123 nodes, representing NUTS3 regions in Europe and 318 edges between them, with an average degree of 5.2. Edges are present between two regions if they are connected by at least one inventor, who patents in the given region. The network is again static as it was aggregated in time between the 1990 – 2010 period. On the regional level, the network is a giant component, meaning that any two nodes can be connected through a path, with an average of 3.8 path length. The diameter of the graph is 9, which is the longest path between two nodes.

The most prominent observation on Figure 11 is that there are no observations from three of the EU8 countries. Lithuania, Latvia have no nodes on the map. This means that no knowledge spillover is taking place in the Baltic countries of the EU8. Inventors who interact with foreigners do not keep on patenting and thus do not pass on the new knowledge learnt from the OECD collaborators.

Figure 11 - Regional collaboration network



Source: Author's own visualization using QGIS

Inventors in various CEE regions have a tendency to collaborate with those in given western European countries. Regarding connectedness Budapest and the nearby Pest region have the largest number of edges, with more than 40 degrees. Pest is linked northwards to several regions in Sweden, France and Germany. At the same time Northern Great Plain region is connected to Italian regions. On the other hand, Ljubljana region has almost 30 degrees and has edges to Spain, Germany and Belgium. Most Western European regions have a tendency to only have a few edges towards the CEE countries, except for metropolitan regions and capitals, that are more likely to have a diversified scientific field and be able to collaborate with more specific fields.

Budapest has both the highest betweenness and closeness centrality, which is not surprising since this region has also the highest number of degrees. This means that they create most bridges between non-connected nodes and are the closest to all other nodes within the network, characteristics that are related to the high degree measure of the region. On the other end of centralities are three regions of the West Pomerania Province in Poland, with a betweenness centrality of 0 and a closeness centrality of 0.13. The low centrality of these regions derives from being connected to other Polish regions and no Western European ones.

In general, we can see that CEE regions are more likely to collaborate with regions that specialize on the same scientific fields, instead of identifying similar productions that could help produce interdisciplinary knowledge. I come to this conclusion when observing different regions in the same country, or two regions that are from different countries but are in close distance proximity to each other. Unless these regions collaborate with a scientific hub, they are more likely to establish connections with differing regions. Hubs on the other hand, such as Turin, Milan, München or Berlin may connect to up to four of the EU8 countries.

5.2. Econometric analysis

Two methods of econometric analyses were carried out to find a causal relationship of EU accession and to explore whether the free movement of people and the European Research Area had significant effects on the quality of patents produced in the EU8.

5.2.1. Difference-in-differences

When the propensity-score weighted difference-in-differences model was carried out on the CEE regions, from the original 57 control regions only 51 remain in the sample, and from the

145 treatment regions 77 are included in the model. Table 3 shows the number of observations included before and after treatment for both control and treatment groups. The number of observations after matching is low, but more importantly the number of treated observations is higher than the regions in the control group, which might hinder the matching of regions.

Table 3 - Size of groups in the DID

	Baseline	Follow-up	
Control	51	49	100
Treated	77	76	153
	128	125	

Source: Author's own calculation using STATA

The model compares treatment to control groups before and after treatment and then calculates the difference of the two groups after 2004 when previous difference of the two groups is taken into account. For the baseline period Table 4 shows a significant negative difference, which means that regions in the EU8 countries produce patents of worse quality than the regions of the control group. This is no surprise, since on Figure 7, I already observed that the EU8 countries have a slightly lower radicalness level than Romania and Bulgaria. The difference is around 0.02 points on the figure, and when covariates were introduced for the matching this difference increases to 0.147. The increase in difference between the control and treatment group is the result of the kernel-based propensity matching, which helped ensure causality, and that I compared regions with similar characteristics to each other.

In the second period, after the treatment a positive but non-significant difference was observed after matching on Table 4. While the difference between treatment and control is not significant, if we exclude the trend learnt from the control group a statistically significant difference-in-differences result is gained. A slight difference was found, that confirms the hypotheses, that entering the EU in 2004 had a positive effect on the radicalness of patents. This means that those regions, that were part of the accession to the EU had an 18.3% higher radicalness measure, than regions in the non-EU control group. Since I use patents as measures for knowledge flows, I can conclude that entering the European Union allowed the EU8 countries to significantly improve their patents, showing for an increased knowledge that may be attributed to foreign inventors coming to EU8 regions as well as CEE inventors leaving abroad and then returning with new knowledge. This relationship is not causal: accession to the EU has other advantages besides becoming part of the European Research Area and allowing free movement of people. At the same time transaction costs are reduced, benefits and investments are flowing in to the countries and international recognition is improving as well.

Table 4 - DID Output

Outcome	**Radicalness**	**Stand.Error**	**T**	**P> |t|**
Baseline Diff (T – C)	-0.147	0.048	-3.09	0.002 ***
Follow-up Diff (T – C)	0.036	0.050	0.71	0.47
DID	0.183	0.069	2.65	0.009 ***

Source: Author's own calculation using STATA

5.2.2. OLS

I carried out two OLS regressions. The first where I explored the effect of the E-I network indicator, treated regions (regions of the EU8), and period (where 1 is period after the 2004 treatment, and 1 is before) on average radicalness of patents within a region. In the second model I included all the variables, which were used in the difference-in-differences method as covariates for the propensity score matching. These are number of patents in the region, number of inventors in the region, area of the region and population density.

Table 5 - OLS Output

	Effect of E-I on Radicalness	
Variables	1	2
Treated	0.0609*** (0.0205)	0.0569** (0.0229)
Period	0.0386** (0.0184)	0.0415** (0.0201)
Patent number		0.0002 (0.0002)
Inventor number		-0.0001 (0.0001)
Region area		~ 0.0 (~ 0.0)
Population density		~ 0.0 (~ 0.0)
E-I index	-0.0431** (0.0193)	-0.0358 (0.0224)
Constant	0.242*** (0.0217)	0.230*** (0.0306)
R^2	0.049	0.061
N	380	330

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Source: Author's own calculation using STATA

The OLS regressions provided further understanding and proof of how accession to the EU helped the development of innovative activity in the EU8 regions. As presented in Table 5, in both models I found a statistically significant positive effect on regions that were in the EU8 countries, and on time. The positive effect of time is evidence, that there is a general positive trend of radicalness as was seen on Figure 4. In Model 1 I only explored the effect of the E-I indicator without regional control variables. The E-I index has a negative significant effect on average radicalness on the 5% significance level. The higher the E – I index, the less radical regional patenting is on average. This means that regions that were more outward looking were worse at innovative patenting.

I found that when regional variables were introduced, the E-I index became statistically non-significant but still has a negative effect on radicalness. In Model 2 it is also interesting to note that neither the number of patents, nor the number of inventors in the region effect radicalness of patents significantly. And while it is not a statistically significant result, it is even more surprising that the higher the number of inventors, the less radical the average patents are within the region. Model 1 has a lower R-squared compared to Model 2, that could mean that the relationship between the dependent variables and the linear model is less strong, but the higher the number of variables the stronger the model gets. Thus, there is no noteworthy gain of explanatory power of the second model, when including the regional control variables.

6. Conclusion and policy recommendation

In this thesis I sought to show how the 2004 European Union enlargement effected the radicalness of patents produced in the EU8 countries. The European Research Area was established in 2000 in order to boost international and interregional inventor collaboration, advance Research and Development and increase innovation. To study this, I used the 2015 OECD REGPAT database, merged with the 2013 OECD database on patent quality and 2013 database on EUROSTAT NUTS3 regional level data. I explored the question with social network and econometric analysis. An inventor and a regional level collaboration network were created to explore the position of international inventors and the possible knowledge flows. A propensity score weighted difference-in-differences and an OLS model were also created to explore causal relations between EU acceptance, regional factors and the type of links regions have (more internal or external) on the average radicalness of the regions.

I found that the position of the OECD inventors in the collaboration network is marginal. Not only is the number of foreign inventors low within the connected components, they are also on the periphery. When their role as brokers was explored, I found that on average they are not crucial in connecting clusters to each other. They are also less highly connected on average, as can be seen when comparing their degree centrality to those of CEE inventors. The regional collaboration network showed that only in a few instances can a knowledge transfer from an OECD inventor have a spill-over effect as inventors do not keep on patenting after collaborating with a foreigner. The difference-in-differences model found a significant positive increase in the radicalness of patents in the EU8 countries after joining the EU. The difference - after controlling for the general trend, that is being observed in the control group – is 18 percentage points after 2004. When the E-I index is included I found that the increased regional patenting radicalness is not an effect of regions establishing new, outside links, but instead inward-looking regions are more successful in patenting.

The policy implications of my findings are clear for the European Research Area. The aim of increasing collaboration is novel and important as I clearly showed that the radicalness of patents produced in collaboration are higher compared to patents created by only CEE inventors. At the same time, I also pointed out that simply reducing the costs of collaboration is not enough. The position of international inventors also matters if a long-term positive knowledge spillover is expected. Foreign inventors need to be imbedded in the network, work on patents with a large number of inventors, or stay longer and collaborate with various institutions, and on numerous patents.

Secondly, I saw that inward-looking regions are more successful in terms of radical patenting. This is because inventors establish links among each other and keep on working and evolving together. The result is that hubs of innovation are created that are not improved when outside connections are included. This raises the question of whether it makes sense to simultaneously increase the number of collaborations and try to close the gap between metropolitan and urban regions. Introducing foreign inventors to urban regions might not result in a knowledge spillover to more metropolitan regions, as these regions have little intention to establish connections with less developed institutions.

Even if joining the European Union had a clear positive effect on the EU8 countries, this effect cannot be regarded to the European policy for research and technological development. The radicalness of patents created in international collaboration is higher than of those created in the CEE region, and joining the EU had an 18-percentage point increase effect on average radicalness in the EU8. On the other hand, I was not able to find a boosting quality effect that was a result of the financial help of the Framework Programmes and the openness of the European Research Area. While Arrieta, Pammolli and Petersen (2017) showed a hypothetical proof for the failure of the ERA, I showed that the increase in patent radicalness is due to the institutional advantages of joining the EU, not due to increased knowledge flow to the region.

There are limitations to this thesis, the most important one being that the study is not repeatable as the 2004 EU accession was a one-time procedure that cannot be reproduced. Data availability and incorrect data gathering are issues that anyone dealing with data analysis needs to be aware of. Due to the law of large numbers, and the extension of the data, this will not influence my results significantly. Lastly, this thesis carries out an overall, generalized study that is backed up by data only. Innovative firms and inventors are not personally identified and interviewed. Such interviews would allow me to find further support for or question the results.

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. https://doi.org/10.1016/S0048-7333(01)00184-6

Anselin, L., Varga, A., & Acs, Z. (1997). Local Geographic Spillovers Between University Research and High Technology Innovations. *Journal of Urban Economics, 42*, 422–448.

Archibugi, D., & Michie, J. (1995). *The Globalisation of Technology: A New Taxonomy. 19*. https://doi.org/10.1093/oxfordjournals.cje.a035299

Balland, P.-A., Belso-Martínez, J. A., & Morrison, A. (2016). The Dynamics of Technical and Business Knowledge Networks in Industrial Clusters: Embeddedness, Status, or Proximity? *Economic Geography*, *92*(1), 35–60. https://doi.org/10.1080/00130095.2015.1094370

Balland, P.-A., & Rigby, D. (2017). The Geography of Complex Knowledge. *Economic Geography*, *93*(1), 1–23. https://doi.org/10.1080/00130095.2016.1205947

Bathelt, H., Malmberg, A., & Maskell, P. (2004). Clusters and knowledge: local buzz, global pipelines and the process of knowledge creation. *Progress in Human Geography*, *28*(1), 31–56. https://doi.org/10.1191/0309132504ph469oa

Beaudry, C., & Schiffauerova, A. (2011). Impacts of collaboration and network indicators on patent quality: The case of Canadian nanotechnology innovation. *European Management Journal*, *29*(5), 362–376. https://doi.org/10.1016/j.emj.2011.03.001

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

Boschma, R., & Fornahl, D. (2011). Cluster Evolution and a Roadmap for Future Research. *Regional Studies*, *45*(10), 1295–1298. https://doi.org/10.1080/00343404.2011.633253

Brandmüller, T., Önnerfors, Å., Reinecke, P., & Statistical Office of the European Communities. (2018). *Eurostat regional yearbook: 2018 edition.* Retrieved from http://dx.publications.europa.eu/10.2785/231975

Breschi, S., & Lissoni, F. (2001). Knowledge Spillovers and Local Innovation Systems: A Critical Survey. *Industrial and Corporate Change*, *10*(4), 975–1005. https://doi.org/10.1093/icc/10.4.975

Chessa, A., Morescalchi, A., Pammolli, F., Penner, O., Petersen, A. M., & Riccaboni, M. (2013). Is Europe Evolving Toward an Integrated Research Area? *Science*, *339*(6120), 650–651. https://doi.org/10.1126/science.1227970

Commission of the European Communities. (1997, May 11). *Fifth Framework Programme for Research and Technological Development*.

Commission of the European Communities. (2000). *Towards a European research area*. 38.

Desrochers, P. (1998). On the abuse of patents as economic indicators. *The Quarterly Journal of Austrian Economics*, *1*(4), 51. https://doi.org/10.1007/s12113-998-1026-z

Doria Arrieta, O. A., Pammolli, F., & Petersen, A. M. (2017). Quantifying the negative impact of brain drain on the integration of European science. *Science Advances*, *3*(4), e1602232. https://doi.org/10.1126/sciadv.1602232

Fitjar, R. D., & Huber, F. (2015). Global pipelines for innovation: insights from the case of Norway. *Journal of Economic Geography*, *15*(3), 561–583. https://doi.org/10.1093/jeg/lbu017

Giuliani, E. (2013). Network dynamics in regional clusters: Evidence from Chile. *Research Policy*, *42*(8), 1406–1419. https://doi.org/10.1016/j.respol.2013.04.002

Glückler, J. (2007). Economic geography and the evolution of networks. *Journal of Economic Geography*, *7*(5), 619–634. https://doi.org/10.1093/jeg/lbm023

Griliches, Z. (1990). Patent Statistics as Economic Indicators: A Survey. *Journal of Economic Literature*, *28*(4), 1661–1707. Retrieved from JSTOR.
Grillitsch, M., & Nilsson, M. (2015). Innovation in peripheral regions: Do collaborations compensate for a lack of local knowledge spillovers? *The Annals of Regional Science*, *54*. https://doi.org/10.1007/s00168-014-0655-8
Grossmann, V., & Stadelmann, D. (2011). Does International Mobility of High-Skilled Workers Aggravate Between-Country Inequality? *Journal of Development Economics*, *95*, 88–94. https://doi.org/10.1016/j.jdeveco.2010.04.007
Guan, J., & Chen, Z. (2012). Patent collaboration and international knowledge flow. *Information Processing & Management*, *48*(1), 170–181. https://doi.org/10.1016/j.ipm.2011.03.001
Guellec, D., & van Pottelsberghe de la Potterie, B. (2001). The internationalisation of technology analysed with patent data. *Research Policy*, *30*(8), 1253–1266. https://doi.org/10.1016/S0048-7333(00)00149-9
Hall, B., Jaffe, A., & Trajtenberg, M. (2001). The NBER Patent Citations Data File: Lessons, Insights and Methodological Tools. *Tel Aviv, Papers*, *8498*.
Hoekman, J., Frenken, K., & van Oort, F. (2009). The geography of collaborative knowledge production in Europe. *The Annals of Regional Science*, *43*(3), 721–738. https://doi.org/10.1007/s00168-008-0252-9
Jaffe, A. B., Trajtenberg, M., & Henderson, R. (1993). Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *The Quarterly Journal of Economics*, *108*(3), 577–598. https://doi.org/10.2307/2118401
Juhász, S., & Lengyel, B. (2018). Creation and persistence of ties in cluster knowledge networks. *Journal of Economic Geography*, *18*(6), 1203–1226. https://doi.org/10.1093/jeg/lbx039
Krackhardt, D., & Stern, R. N. (1988). Informal Networks and Organizational Crises: An Experimental Simulation. *Social Psychology Quarterly*, *51*(2), 123–140. https://doi.org/10.2307/2786835
Lengyel, B., & Leskó, M. (2016). International Collaboration and Spatial Dynamics of US Patenting in Central and Eastern Europe 1981-2010. *PLOS ONE*, *11*(11), e0166034. https://doi.org/10.1371/journal.pone.0166034
Marzucchi, A., Antonioli, D., & Montresor, S. (2015). Industry–research co-operation within and across regional boundaries. What does innovation policy add? *Papers in Regional Science*, *94*(3), 499–524. https://doi.org/10.1111/pirs.12079
Montobbio, F., & Sterzi, V. (2013). The Globalization of Technology in Emerging Markets: A Gravity Model on the Determinants of International Patent Collaborations. *World Development*, *44*, 281–299. https://doi.org/10.1016/j.worlddev.2012.11.017
Morescalchi, A., Pammolli, F., Penner, O., M, P. A., & Riccaboni, M. (2013). *Networks of innovators within and across borders. Evidence from patent data* (No. 4/2013). Retrieved from IMT Institute for Advanced Studies Lucca website: https://ideas.repec.org/p/ial/wpaper/4-2013.html
O Lanjouw, J., Pakes, A., & Putnam, J. (1998). How to Count Patents and Value Intellectual Property: The Uses of Patent Renewal and Application Data. *Journal of Industrial Economics*, *46*, 405–432. https://doi.org/10.1111/1467-6451.00081
Penrose, E. (1973). International Patenting and the Less-Developed Countries. *The Economic Journal*, *83*(331), 768–786. https://doi.org/10.2307/2230670
Petersen, A. M., & Puliga, M. (2017). High-skilled labour mobility in Europe before and after the 2004 enlargement. *Journal of the Royal Society, Interface*, *14*(128). https://doi.org/10.1098/rsif.2017.0030

Picci, L. (2010). The internationalization of inventive activity: A gravity model using patent data. *Research Policy*, *39*(8), 1070–1081.
Policy for research and technological development | Fact Sheets on the European Union | European Parliament. (n.d.). Retrieved May 28, 2019, from http://www.europarl.europa.eu/factsheets/en/sheet/66/policy-for-research-and-technological-development
Squicciarini, M., Dernis, H., & Criscuolo, C. (2013). *Measuring Patent Quality*. https://doi.org/10.1787/5k4522wkw1r8-en
Storper, M., & Venables, A. J. (2004). Buzz: face-to-face contact and the urban economy. *Journal of Economic Geography*, *4*(4), 351–370. https://doi.org/10.1093/jnlecg/lbh027
Varga, A., & Schalk, H. (2004). Knowledge Spillovers, Agglomeration and Macroeconomic Growth: An Empirical Approach. *Regional Studies*, *38*(8), 977–989. https://doi.org/10.1080/0034340042000280974
Varga, A., & Sebestyén, T. (2017). Does EU Framework Program Participation Affect Regional Innovation? The Differentiating Role of Economic Development. *International Regional Science Review*, *40*(4), 405–439. https://doi.org/10.1177/0160017616642821
Wagner, C. S., Park, H. W., & Leydesdorff, L. (2015). The Continuing Growth of Global Cooperation Networks in Research: A Conundrum for National Governments. *PLOS ONE*, *10*(7), e0131816. https://doi.org/10.1371/journal.pone.0131816
Wal, A. L. J. T., & Boschma, R. (2011). Co-evolution of Firms, Industries and Networks in Space. *Regional Studies*, *45*(7), 919–933. https://doi.org/10.1080/00343400802662658