Utrecht University

Exploring the Link between Airport Connectivity and Knowledge Complexity

A European case study

Jasper Ott (5650631): j.ott@students.uu.nl
Orlando Vazquez (6674828) : o.m.vazquezvillegas@students.uu.nl
Siting Wu (6422144): s.wu4@students.uu.nl
Anh Qui Huynh (6901018) : a.q.huynh@students.uu.nl
Lukas Schubotz (6674054): l.m.schubotz@students.uu.nl

Phone number: +31651133467

Word count: 7587

Table of contents

1 Introduction

During the last three decades, many of the European economies have encountered an unsatisfactory growth due to declining productivity and competitiveness (Veugelers, 2017; Pradhan et al., 2019). To stimulate regional economic growth and enhance productivity, increasing attention has been paid to stress the importance of knowledge complexity on regional economic development (Hidalgo & Hausmann, 2009; Balland & Rigby, 2017). A region with a wide range of knowledge or capabilities is more likely to create unique, complex and diverse products, and thereby tends to sustain a competitive advantage and experience long-run economic growth (Hidalgo & Hausmann, 2009; Balland et al., 2019). Balland and Rigby (2017) proved this argument by introducing a knowledge complexity index (KCI) using network-based techniques based on patent data to quantify knowledge production capability across US cities. They suggested that regions should be encouraged to develop technologies that are more complex than their current knowledge bases to strengthen global competitiveness (Balland & Rigby, 2017).

Understanding the determinants of knowledge complexity is essential for facilitating the transformation of regional knowledge structure towards a high complexity level across European countries. Several studies have already addressed the determinants of knowledge complexity. For example, Hidalgo and Hausmann (2009) pointed out that the complexity of products and countries are strongly correlated with national income per capita. Additionally, Zhu and Li (2016) took the effect of human capital into account, indicating that the educational level is closely associated with the development of complexity. Another study showed a positive correlation between scientific productivity and the development of knowledge complexity (Laverde-Rojas and Correa, 2019).

Although previous studies have exhibited some of the contributions of economic and human capital on the development of complexity, the impact of the connectivity between regions, which could have enormous potential in facilitating knowledge exchange and shaping knowledge production, has not yet been studied systematically. Cantwell and Salmon (2018) investigated the development of knowledge complexity in the globalization context and concluded that the diverse knowledge fields and distinct geographic locations are conducive to increase regional knowledge complexity. They further highlighted the value of connecting previously unconnected regions on promoting innovative knowledge production (Cantwell & Salmon, 2018). Furthermore, various modes of transport infrastructure have been proved to be the facilitation of creating connectivity, among which air transport is significantly important in the cross-countries context (Shepherd, 2016; Kararach, 2014). The positive relationship between connectivity and knowledge complexity has been theoretically addressed based on simplistic theoretical models (van Zon & Mupela, 2016). However, empirical studies on the relationship between connectivity, more specifically by air transport, and knowledge complexity are lacking.

To provide empirical study on the influence of airport connectivity on the regional knowledge complexity, our study is designed to answer the following question: To what extent does airport connectivity contribute to the development of regional knowledge complexity in Europe?

By understanding the underlying principles of knowledge complexity on a NUTS2 level, we are able to provide valuable policy recommendations to the European Union (EU). These recommendations can advise the EU on which measures to take in order to improve the complexity of knowledge, specifically for regions that lack behind in terms of knowledge complexity. Current aviation funds have mainly been allocated to Spain, Greece, Poland and Italy, collectively receiving 75 percent of the total amount of funding (European Court of Auditors, 2014). Whereas, based on the findings in this report, it makes sense to allocate funding to Greece and Poland based on the low KCI values of specific Polish and Greek regions, the large funds allocated to Italy and Spain could be allocated more efficiently. Specifically, to the regions that we identified to be lacking behind in terms of knowledge complexity: Romania, Ukraine, Hungary and Bulgaria. Specific targeting of aviation funds can strongly contribute to minimization of economic inequality between EU countries since knowledge complexity is known to have a strong contribution to economic growth (Hidalgo & Hausmann, 2009; Balland & Rigby, 2017).

The research is organized into three parts: Firstly, we calculate the knowledge complexity of NUTS2-regions in European countries using the methods developed by Balland and Rigby (2017). Following this we use a social network analysis (SNA) to quantify the airport connectivity of each of the NUTS2-regions. Lastly, we aim to address the research gap by investigating the extent to which connectivity contributes to the development of regional knowledge complexity.

2 Theoretical framework

In this section we will discuss the principle of knowledge complexity in more detail and elaborate on its relation to airport connectivity. Furthermore, we will identify and justify the different control variables that will be taken into consideration in this study.

2.1 Knowledge Complexity

Knowledge complexity can be defined as a function of the required components and their interdependencies to generate this kind of knowledge (Fleming & Sorenson, 2001). Complex knowledge is mainly dependent on tacit knowledge, a form of non-codified knowledge that is embedded in people's minds and therefore harder to diffuse (Kogut and Zander, 1993). Knowledge bases of regions differ both in their technological composition and value (Boschma, 2005). The value of knowledge bases depends on the complexity of the embedded knowledge.

The variety of knowledge that a country or region possesses, but also the extent to which they are able to combine this knowledge, strongly determines the amount of capabilities within that country or region (Hausmann et al., 2013). Countries or regions that combine their capabilities in an efficient and unique way, are able to develop new technologies with greater complexity than regions who are not able to do so (Hidalgo & Hausmann, 2009). In other words, knowledge complexity can be explained by two dimensions of technological profiles - *diversity* and *ubiquity*, where diversity refers to the capabilities of regions that possess a diverse range of technologies, and ubiquity suggests whether the regions are capable to develop unique and complex technologies. High diversity and low ubiquity contributes to a more complex knowledge structure of the region (Balland et al., 2019).

That complex knowledge indeed contributes to economic growth has been demonstrated previously (Hidalgo & Hausmann, 2009). Hidalgo & Hausmann created the Index of Economic Complexity, where economic complexity is defined as a measure of the knowledge in a society as expressed in the products it makes. In this study, they showed that the Index of Economic Complexity beats competitiveness measures such as the World Economic Forum's Global Competitiveness Index by a tenfold in predicting economic growth for the following decade (Hidalgo & Hausmann, 2009). These results underline the strong positive impact of knowledge complexity on economic growth and therefore the importance of knowledge complexity to regions or countries.

2.2 Connectivity and knowledge complexity

In the previous section, we defined knowledge complexity which consists of technological diversity and ubiquity dimensions, and identified the relationship between knowledge complexity and economic growth. In order to address our research question, we will consider the relationship between connectivity and knowledge complexity.

Numerous studies already identified geographical proximity as a key determinant of knowledge creation (Allen, 1977; Glaeser, Kallal, Scheinkman, and Shleifer, 2000). The underlying principle being that geographical proximity eases face-to-face interaction within networks which enables the transfer of tacit knowledge between individuals or organizations (Conventz, 2014). In turn, the transfer of tacit knowledge is known to strongly contribute to the creation of complex knowledge (Grant, 1996). From this theory we can deduct that geographical proximity within a network stimulates the creation of complex knowledge.

One way of increasing the “perceived” geographical proximity within a network is by improving the connectivity between its actors (Coventz, 2014). We view perceived geographical proximity as the effort it takes to travel the geographical distance between two actors, contrary to the classical definition of geographical proximity as the inverse of the linear distance between two actors (Boschma, 2005). An increase in connectivity can be established by strengthening the infrastructure between the actors of the network. Aviation is considered as the most efficient way since it allows for the greatest reduction of travel time, cost and effort, especially on large distances, which are the main determinants of perceived geographical proximity (Bernela et al., 2019).

There are several measures that can indicate the centrality of an airport. In this report the centrality measures degree and betweenness centrality will be used. Degree centrality is the most straightforward indicator in analyzing airport connectivity since it shows the number of direct connections linked to the airports and therefore gives an indication of the overall connectedness of an airport (Rocha, 2017). Betweenness centrality is another important indicator characterizing how influential an airport is within the network by calculating the extent to which a node lies on paths between other nodes. Airports with high betweenness may have considerable influence within the network by virtue of their control over passengers passing between others (Goh et al., 2003). Analyzing both indicators resulted in a solid estimation of an airports’ centrality within the network.

Now we identified our measures of connectivity, we can combine the principle that the creation of complex knowledge depends on geographical proximity and the principle that air transport can increase perceived geographical proximity to formulate the following two hypotheses:

H1: Regions with airports characterized by high degree centrality possess higher levels of knowledge complexity.

H2: Regions with airports characterized by high betweenness centrality possess higher levels of knowledge complexity.

2.3 Control variables: socio-economic factors

In the previous section, we identified the relationship between knowledge complexity and connectivity. However, there are also other factors that are known to have an impact on both the knowledge complexity and the airport connectivity of a region. These factors will serve as our control variables, which we will discuss in this section.

The first control variable that will be used is the share of employees in technology and knowledge-intensive sectors within a region. Research has indicated that the number of high-tech employees positively influences the innovativeness of a country, indicated by the amount of patents within that country (Baesu et al., 2015). The creation of a patentable product goes hand-in-hand with the creation of new knowledge, which is spilled-over as well as a result of the publication of the patent (Acs & Sanders, 2011). Therefore, we would argue that regions having a high number of high-tech employees will also create more knowledge, which leads to a higher overall knowledge complexity within that region and vice versa. According to the theory of proximity and in relation to airport connectivity, a region that has a high density of high-tech employees is attractive for other regions to connect to, since this would allow them to get access to the tacit knowledge that is accumulated in those individuals (Grant, 1996). As we mentioned earlier, the most efficient way to improve connectivity is by strengthening the air transport between regions (Bernela et al., 2019). Based on these findings, we would argue that high-tech employees are better able to exchange tacit knowledge when they are connected to other regions that is characterized by a high density of high-tech employees. This would in turn contribute to the production of more complex knowledge. Consequently, we expect that a high density of high-tech employees increases the airport connectivity of a region.

The second control variable that we will be taken into consideration is population density. Previous studies have demonstrated that larger population possesses more diverse knowledge (Kline & Boyd, 2010), which can also be combined more easily, resulting in a higher overall knowledge complexity. Furthermore, aviation demand increases with population size (European Union, 2017). As a result, flight going in and out of the region will increase, which has a direct positive effect on the airports degree centrality (European Union, 2017). Additionally, when the number of flights from and to the airport increases, it is likely that the airport will also play a more central role in the aviation network, having a positive effect on the airport's betweenness.

Lastly, the Gross Domestic Product (GDP) per capita will be used as a control variable in order to summarize the economic position of NUTS2-regions in Europe. Furthermore, Hidalgo and Hausmann (2009) have demonstrated that a country's income, measured by GDP per capita, is correlated with the knowledge embedded in economic complexity. On the other hand, airport connectivity facilitates the exchange of tacit knowledge, which may lead to an increase in knowledge complexity. The relationship to both airport connectivity and knowledge complexity makes GDP per capita a solid control variable.

3 Methodology

This chapter presents the data and methods used in this study to investigate the relationship between regional knowledge complexity and airport connectivity.

3.1 Research design

The research design of this study is based on quantitative, deductive methods using cross-sectional data derived from the European Patent Office's Worldwide Statistical Patent Database, the OpenFlights Airports Database and the Eurostat Dissemination Database. By using cross-sectional data, we are able to analyze the relationship between the development of regional knowledge complexity and airport connectivity at a given period of time, taking into account socio-economic factors as control variables.

3.2 Collection of secondary data sources

Our sampling strategy was to utilize secondary data because the research objects consisted of different factors at the NUTS2-level within the EU. Due to the volume and width - not to mention the difficulty of collecting data - secondary data enabled us to save time and effort on procurement, and instead to invest those resources in the data analysis. Since we strived to take the most recent and available data into account - and considering a time lag in patents due to the fact that the application procedure of patents takes typically around three years (EPO, 2014; USPTO, 2019) - we chose 2016 as the investigation year.

As mentioned above, we used the knowledge complexity index conducted in the study of Balland and Rigby (2017) as a guide to build our own knowledge complexity index. For the independent variables, we drew our data from the OpenFlights Airports Database containing all airports and flights available in Europe. From this database, we gathered meaningful variables for our data analysis serving as the basis to calculate airport connectivity. All variables derived from the OpenFlights Airport Database are listed in table 1. Lastly, the control variables were retrieved from the Eurostat database on the NUTS2-level.

Table 1: Variables of the OpenFlights Airports Database

	Variable	**Description**
1	IATA	Three-letter geocode designating individual airports defined by the International Air Transport Association (IATA)
2	Source airports	IATA code of the source airport
3	Destination airports	IATA code of the destination airport
4	Country	Country or territory where airport is located

3.3 Operationalization

3.3.1 Dependent variable: knowledge complexity

It is common to use patent data to measure innovation intermediate or output due to its novelty nature and the clear classifications by technological groups (Castaldi et al., 2015). Previous studies have used single or composite patent indicator(s) such as patent counts, citation, family size, renewals and litigation to evaluate regional technology and knowledge performance (Trajtenberg, 1990; Scherngell & Jansenberger, 2006; Czarnitzki et al., 2009; Castaldi et al., 2015). As already mentioned above, Balland and Rigby (2017) introduced a knowledge complexity index using patent data to quantify knowledge creation. We employed this method in this study to calculate and compare the knowledge complexity in the NUTS2-regions.

To construct the knowledge complexity index for the EU regions, the NUTS-2 regions are considered as the producers of particular technologies. We classified patent data into five overarching technological fields and in total 35 technological sub-fields according to the classification proposed by Schmoch (2008). We further considered a region to have a relative technological advantage (RTA) in a specific technological field, if the share of patent count in technology i is higher than the share of technology i in the entire EU. Having calculated the RTA of every NUTS-2 region in the 35 technological sub-fields, we constructed a region-tech network (two-mode network) connecting regions to the technological sub-fields in which they have RTA, to compute the KCI. The network was build upon an $i \times r$ two-mode adjacency matrix (denoted by M), where $i = 35$ are the technological sub-fields and $r = 298$ the NUTS-2 regions. We calculated the degree centrality of each of the NUTS-2 region ($K_{r,0}$) as the number of technological sub-fields that a region has an RTA in (1), and the degree centrality of each technological sub-field ($K_{i,0}$) is presented as the aggregative numbers of regions that have RTA in this technological sub-field (2).

$$K_{r,0} = \sum_i M_{r,i} \quad (1)$$

$$K_{i,0} = \sum_r M_{r,i} \quad (2)$$

Following the method developed by Hidalgo & Hausmann (2009) and Balland & Rigby (2017), the calculation of knowledge complexity index was achieved by a technique called method of reflection combining the diversity and ubiquity of both NUTS-2 regions and technological sub-fields. Following from (1) and (2), the average ubiquity of the technological profile in which a NUTS-2 region has RTA is given by $K_{r,1}$(3), whereas the diversity of a region that has RTA in a specific technological sub-field i is given by $K_{i,1}$(4).

$$K_{r,1} = \frac{1}{K_{r,0}} \sum_i M_{r,i} K_{i,0} \quad (3)$$

$$K_{i,1} = \frac{1}{K_{i,0}} \sum_i M_{r,i} K_{r,0} \quad (4)$$

Continuing the method of reflection to the next layer, the average diversity of the technological profile in which a NUTS-2 region has RTA is given by $K_{r,2}$(5), whereas the ubiquity of a region that has RTA in a specific technological sub-field i is given by $K_{i,2}$(6).

$$K_{r,2} = \frac{1}{K_{r,0}} \sum_i \quad M_{r,i} \, K_{i,1} \tag{5}$$

$$K_{i,2} = \frac{1}{K_{i,0}} \sum_i \quad M_{r,i} \, K_{r,1} \tag{6}$$

The KCI of each NUTS-2 region and technological sub-field can then be computed through combining the diversity and ubiquity measures. In this study, our only interest lies in the knowledge complexity in the region. Since a high technological diversity and low ubiquity contribute to a higher knowledge complexity, the KCI of NUTS-2 regions was computed by using diversity divided by ubiquity (7).

$$KCI_{regions} = \frac{K_{r,2}}{K_{r,1}} \tag{7}$$

3.3.2 Independent variable: airport connectivity

In order to measure the airport connectivity we performed a social network analysis (SNA) using a directed graph, in which airports represent nodes and the outgoing flights are the edges connecting them. The purpose of the SNA was to measure the airport connectivity based on two centrality measures: degree centrality and betweenness centrality. Both centrality measures form the basis for the OLS regression analysis by defining the independent variables of airports' connectivity.

For the computation of the degree centrality of airports, a directed graph based on an one-mode adjacency matrix of source airports and destination airports considering only outdegrees (outgoing flights) was multiplied with a unit vector of 1. Therefore, the degree centrality identifies highly central airports based on the number of outdegrees, where each outdegree represents a connecting flight to another airport. An example on how to compute the degree centrality can be seen in figure 1.

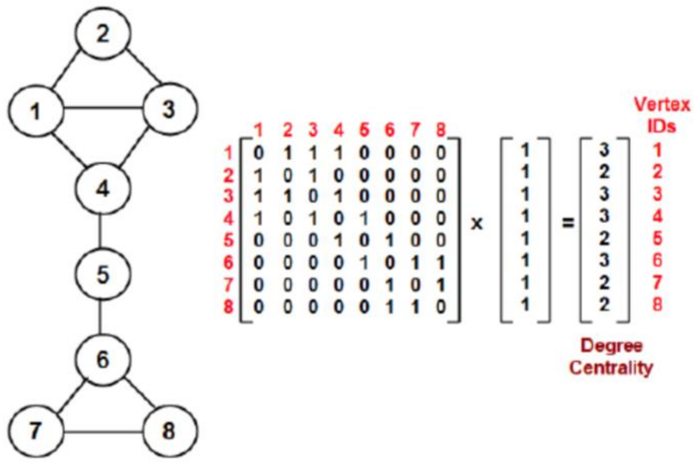


Figure 1: Illustration of the Computation of the Degree Centrality (Meghanathan, 2016, p. 9)

The betweenness centrality, on the other hand, is defined by the number of geodesics (shortest paths) passing through a vertex and measures the extent to which a node lies on paths between any other node (Song and Yeo, 2017). To determine the betweenness centrality, we followed two steps of according to the algorithm of Brandes (2001). First, the length (sum of weights of a vertex's edges) and number of the shortest paths between all pairs were calculated, which results in the pair-dependency. The pair-dependency is denoted by δ_{st}, where *s* and *t* stands for the minimum distance between the airport *s* and *t*. Second, all pair dependencies were summed up. Both steps to compute the betweenness centrality can be summarized by the following equation:

$$C_B(v) = \sum_{s \neq v \neq t \in V} \delta_{st}(v)$$

The SNA results for airport connectivity were based on a network containing 853 nodes representing airports and 11689 edges illustrating connecting flights between all airports within the European Union. The descriptive statistics of airports' degree centrality and betweenness are shown in table 2.

Table 2: Descriptive statistics of centrality measures

	Degree Centrality	Betweenness
Maximum Value	141	48875.14
Minimum Value	0	0
Mean	13.70	871.11
Median	4	0

Standard Deviation	22.87	3637.06

Table 2 shows that the spectrum of airport connectivity is broad in terms of both centrality measures. This can be derived from the range between the minimum and maximum, from the difference between the mean and median, and from the standard deviation. Given the fact that the mean of both centrality measures is greater than the median, it can be also concluded that the value distribution of the degree and betweenness centrality is right-skewed. With regard to the degree centrality, this means that a relatively large number of airports operates a low number of routes to connect to other airports, whereas a few airports possess a significant higher number of routes. In the view of the betweenness centrality, this suggests that a relatively large number of airports are not on the shortest paths between other airports, while a few airports have a significant higher number of shortest paths passing through their vertex. Furthermore, the standard deviation of both centrality measures stresses out that the values are not close to the mean but are distributed unevenly. This result can also be seen in the histogram of degree and betweenness centrality (S1 and S2).

Based on these results, we incorporated the natural logarithms of airports' degree and betweenness centrality in all regression models to deal with their skewed distribution. Furthermore, we detected multicollinearity (VIF > 5) between the natural logarithm of both centrality measures, if they were included together in one regression model. For this reason, both independent variables were observed separately in the second and third model of each regression table.

3.3.3 Control variables: socio-economic factors

As broadly explained in section 2.3, the share of employees in technology and knowledge-intensive sectors, population density and GDP per capita might affect the development of knowledge complexity and the airport connectivity. Thus, these factors are chosen as control variables to reduce the extraneous effects. Detailed descriptions of control variables are shown in table 3.

Table 3: Operationalization table

	Variable	Category	Indicator	Description
1	Dependent variable	Knowledge complexity	Knowledge complexity index	The diverse and unique knowledge that a NUTS2-region possesses
2	Independent variables	Airport connectivity	Degree centrality	The number of outgoing connection routes to other airports
3			Betweenness centrality	The number of geodesics (shortest paths) passing through an airport
4	Control variables	Socio-economic factors	Share of HTC employees	The share of employees in technology and knowledge-intensive sectors

5			Population density	The number of inhabitants living in each square kilometer
6			GDP per capita	Country’s GDP divided by its total population

3.4 Data analysis

In order to analyze the relationship between the independent and the dependent variables, we perform a multiple OLS regression model. In doing so, we are able to include the degree (DC) and betweenness centrality (BC) of airports as independent variables along with the control variables GDP per capita (GDPpc), population density (Pd), researcher density (Rd) and high-tech employee density (HTEd) in the model in order to test the relationship with the dependent variable knowledge complexity (KC):

$$KC = ß_0 + ß_1 DC + ß_2 BC + ß_3 GDPpc + ß_4 Pd + ß_5 Rd + ß_6 HTEd + e_i$$

3.5 Research quality indicators

The research design in this study is adapted to the EU as its country members are the objects of interest. Still, the underlying research design can also be applied to other countries, such as China or the USA. Therefore, this research design can be used as a guide for further study in other parts of the world.

Regarding the data analysis, we will perform the OLS regression analysis according to the Gauss-Markov theorem. This means that during the analysis, care is taken to ensure that no multicollinearity, exogeneity or heteroscedasticity occurs in the model. To ensure this, the analysis is carried out together with the Breusch-Pagan test and the variance inflation factor in order to identify and counteract potential threats to the model.

However, this research design shows some limitations. First of all, an OLS regression analysis can only test for correlation but generally not for causality. Therefore, we interpret the relationship between the dependent, independent, and control variables in a neutral manner. Secondly, the data set consists of several data sources. To match all the data, we select the datasets carefully in terms of renowned data sources (mainly Eurostat) and complementary time periods, in this case, the year 2016. On this basis, we make sure that this research and the resulting findings are as valid and reliable as possible.

4 Results

In this section we will present our results of our regression models and interpret those results. In order to test our hypotheses, we ran four OLS regression models with a sample of 172 observations that are shown in table 4. The first model investigates the relationship between the dependent variable and the control variable. In the second and third model, we separately included both independent variables – the natural logarithm of airports' degree (*ln_degree*) and betweenness centrality (*ln_betweeness*) – together with all control variables (the share of employees in technology and knowledge-intensive sectors, population density and GDP per capita) in order to test the relation between the dependent variable and independent variables, while controlling for socio-economic factors. The fourth model represents the full model, in which both independent variables and all control variables are included.

Table 4: OLS Regression Models with KCI as the Dependent Variable

	Dependent variable:			
	KCI			
	(1)	(2)	(3)	(4)
ln_degree		0.156		0.190
		(0.118)		(0.247)
ln_betweenness			0.049	-0.015
			(0.046)	(0.095)
Share_of_HTC_employees	0.395***	0.364***	0.371***	0.365***
	(0.079)	(0.082)	(0.082)	(0.083)
Population_density	0.0002	0.0001	0.0001	0.0001
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
GDP_per_capita	0.00000	0.000	-0.00000	0.00000
	(0.00001)	(0.00001)	(0.00001)	(0.00001)
Constant	11.872***	11.613***	11.797***	11.579***
	(0.299)	(0.357)	(0.307)	(0.417)
Observations	172	172	172	172
R^2	0.205	0.213	0.210	0.213
Adjusted R^2	0.191	0.194	0.191	0.189
Residual Std. Error	1.672 (df = 168)	1.668 (df = 167)	1.671 (df = 167)	1.673 (df = 166)
F Statistic	14.416*** (df = 3; 168)	11.294*** (df = 4; 167)	11.112*** (df = 4; 167)	8.988*** (df = 5; 166)

Note:	*p<0.10;**p<0.05;***p<0.01

Table 4 shows that neither the independent variable *ln_degree* nor *ln_betweenness* has a statistically significant relationship to the KCI. This applies to both models 2 and 3, in which the independent variable is treated separately, and to the full model 4. By looking at the control variables, only the *Share_of_HTC_employees* can be significantly and positively associated with the KCI on the 1% level in all four models. All other control variables are not statistically significantly related to the dependent variable. Accordingly, a one-unit increase in the *Share_of_HTC_employees* increases the *KCI* by 0.395 units in model 1, 0.364 units in model 2, 0.371 units in model 3 and 0.365 units in model 4, holding all other variables constant. The adjusted R^2 with a value of 0.191 (model 1), 0.194 (model 2), 0.191 (model 3) and 0.189 (model 4) indicates that the regression model explains the variation of the dependent variable by 19.10%, 19.40%, 19.10% or 18.90%, respectively.

On this basis, both hypotheses 1 and 2 can initially be rejected, implying that regions with airports characterized by a high degree or betweenness centrality might not contribute to a higher level of knowledge complexity. As mentioned in section 2.1, however, KCI is composed of the two sub-dimensions ubiquity and diversity. Therefore, it is plausible that airport connectivity plays distinct roles on these two dimensions. To gain more insights on whether our hypotheses can be accepted or fully rejected, in the following analysis, we looked into the effects of airport connectivity on the two dimensions - ubiquity and diversity - of knowledge complexity, and ran the similar models for these two variables. The results were shown in table 5 and table 6.

Table 5: OLS Regression Models with Ubiquity as the Dependent Variable

	Dependent variable:			
	Ubiquity			
	(5)	(6)	(7)	(8)
ln_degree		-0.592		-0.630
		(0.669)		(1.395)
ln_betweenness			-0.195	0.017
			(0.257)	(0.535)
Share_of_HTC_employees	-2.089***	-1.973***	-1.994***	-1.973***
	(0.446)	(0.466)	(0.464)	(0.467)
Population_density	-0.001	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
GDP_per_capita	0.00002	0.00002	0.00003	0.00002
	(0.0001)	(0.0001)	(0.0001)	(0.0001)

Constant	94.861***	95.845***	95.160***	95.883***
	(1.684)	(2.020)	(1.732)	(2.361)
Observations	172	172	172	172
R^2	0.172	0.176	0.175	0.176
Adjusted R^2	0.158	0.157	0.155	0.151
Residual Std. Error	9.425 (df = 168)	9.431 (df = 167)	9.437 (df = 167)	9.460 (df = 166)
F Statistic	11.666*** (df = 3; 168)	8.933*** (df = 4; 167)	8.871*** (df = 4; 167)	7.104*** (df = 5; 166)
Note:				*p<0.10;**p<0.05;***p<0.01

Regarding the regression table 5, it can be said that both independent variables *ln_degree* and *ln_betweenness* are not statistically significantly associated with the dependent variable *ubiquity*. This result is comparable to the results that are shown in table 4. In terms of control variables, the *Share_of_HTC_employees* variable remains the only regressor in all models that has a significant negative relationship to the dependent variable at the 1% level. The negative coefficients indicate that the ubiquity of the patents produced in a region lower, meaning that the patent they produce are more unique. The adjusted R^2 of all four regression models show that the variation of ubiquity can be explained by 15.80%, if the control variables are taken into account exclusively; by 15.70% and 15.50%, if the independent variables are also included separately in the regression, and by 15.10%, if all variables are taken into account.

Given these results, we can now determine the relationship between the dependent variable *KCI* and independent variables *ln_degree*, *ln_betweenness* as well as the relationship between KCI and the control variable *Share_of_HTC_employees* in a more precise manner. It can be derived that the connectivity of airports in terms of both centrality measures degree and betweenness have no significant relationship with the ubiquity of patents in the corresponding region. Furthermore, the share of employees in technology and knowledge-intensive sectors has a negative impact on the ubiquity of patents in a regions, indicating that the higher the share of employees in technology and knowledge-intensive sectors, the lower the ubiquity of patents in a region, increasing the regions *KCI*.

Table 6: OLS Regression Models with Diversity as the Dependent Variable

	Dependent variable:			
	Diversity			
	(9)	(10)	(11)	(12)
ln_degree		0.057**		0.082*
		(0.023)		(0.048)

ln_betweenness			0.017^{*}	-0.011
			(0.009)	(0.019)
Share_of_HTC_employees	0.063^{***}	0.052^{***}	0.055^{***}	0.052^{***}
	(0.016)	(0.016)	(0.016)	(0.016)
Population_density	0.00004	0.00001	0.00002	0.00001
	(0.00005)	(0.00005)	(0.00005)	(0.00005)
GDP_per_capita	0.00000	0.00000	0.00000	0.00000
	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Constant	11.255^{***}	11.160^{***}	11.229^{***}	11.135^{***}
	(0.060)	(0.070)	(0.061)	(0.082)
Observations	172	172	172	172
R^2	0.173	0.202	0.190	0.204
Adjusted R^2	0.159	0.183	0.171	0.180
Residual Std. Error	0.333 (df = 168)	0.328 (df = 167)	0.331 (df = 167)	0.329 (df = 166)
F Statistic	11.737^{***} (df = 3; 168)	10.587^{***} (df = 4; 167)	9.803^{***} (df = 4; 167)	8.503^{***} (df = 5; 166)
Note:				*p<0.10;**p<0.05;***p<0.01

According to table 6, it can be observed that the natural logarithm of airports' degree centrality *ln_degree* are positively and significantly related to the KCI's sub-dimension *diversity* either if its included as the only independent variable in the model (model 10) but also together with the natural logarithm of airports' betweenness centrality *ln_betweenness* (model 12). This relationship is statistically significant on the 5% level when considered as exclusive regressor, and on the 10% level when incorporated with the variable *ln_betweenness*. Holding all other variables constant, a one-percent increase in ln_degree as an exclusive regressor will change the diversity of patents in technological sub-fields on average by 0.00057 units, whereas in the full model (12) a one-percent increase in ln_degree will change the diversity of patents in technological sub-fields on average by 0.00082 units. Furthermore, the variable *ln_betweenness* shows a statistical and significant positive association with the variable *diversity* only if it is included as an exclusive regressor (model 11), while losing its significant relationship in the full model. The relationship between *ln_betweenness* and *diversity* in model 11 is statistically significant on the 10% level. Correspondingly to the interpretation of ln_degree, a one-percent increase in *ln_betweenness* as exclusive regressor increases the diversity of patents in technological sub-fields on average by 0.00017 units. In the view of these results, we can conclude that the connectivity of airports in terms of degree and betweenness centrality is significantly correlated with the KCI's sub-dimension *diversity.* Accordingly, there is a connection between the KCI and the connectivity of airports through the diversity of patents. As in table 4 and

5, the variable *Share_of_HTC_employees* still remains as the only control variable that has a statistically significant and positive relation to the dependent variable on the 1% level. Hence, the share of employees in technology and knowledge-intensive sectors is related to the *KCI* in that sense that it is positively correlated with the sub-dimension *diversity*, which in turn increases the *KCI*.

5 Discussion

In this study we posed the following research question: *To what extent does airport connectivity contribute to the development of regional knowledge complexity in Europe?* To answer this question, we first calculated a Knowledge Complexity Index (KCI) representing the knowledge complexity of a region. Then, flight data was analyzed to calculate the degree and betweenness centrality of airports within Europe. An Ordinary Least Squares (OLS) regression was performed to quantitatively study the effect of airport connectivity, indicated by the centrality measures betweenness and degree, on the regional knowledge complexity.

Our model has shown that there is no significant relationship between an airports degree and betweenness and the knowledge complexity of the corresponding region. However, once we split our knowledge complexity index into the sub-dimensions *ubiquity* and *diversity* we found that an airports degree and betweenness have a significant positive relation with a regions patent diversity. The effect of degree on knowledge complexity being larger than the effect of betweenness. No significant relationship between an airports degree and betweenness and the ubiquity of patents in a region was found. The results indicate that a large amount of flights, and passengers, going in and out does not increase the KCI of a region, but does increase the patents diversity of the region. As explained previously, a highly connected airport increases the perceived geographical proximity between actors, with geographical proximity being a key determinant of knowledge creation (Allen, 1977; Glaeser, Kallal, Scheinkman, and Shleifer, 2000). The underlying principle is that geographical proximity eases face-to-face interaction within networks which enables the transfer of tacit knowledge between individuals or organizations (Conventz, 2014). In turn, the transfer of tacit knowledge is known to strongly contribute to the creation of complex knowledge (Grant, 1996). Hence, regions with a highly connected airport possess a higher diversity of patents than less connected regions. Furthermore, the patent diversity of a region is dependent on whether this airport is strategically positioned within the aviation network or not, indicated by its betweenness. A possible explanation for this finding is that airports play a substantial role in the economic growth and development of cities and regions. Well connected airports are highly attractive business sites and the agglomeration of firms in the airport surrounding regions can result in a higher patent output of that region, increasing it's patent diversity (Green, 2007).

Striking is the fact that the degree and betweenness of an airport have an effect on the patent diversity of a region and not on its patent ubiquity. These findings indicate that bringing together various forms of knowledge allows for the production of a diverse range of products and corresponding patents, whereas it doesn't allow for specialization in a certain technological field to create unique patents, indicated by the ubiquity measure. The latter might be the result of knowledge bases of regions becoming more similar as a consequence of increasing airport connectivity. When regions connects to each other, their inhabitants share and accumulate knowledge. The result is that knowledge bases of regions start to resemble one another more and more. As a result it might harder to produce unique products and mostly ubiquitous products and corresponding patents will be produced (Sapienza, Parhankangas & Autio, 2004).

5.1 Theoretical implications

Theoretically, this paper adds to the field of regional knowledge complexity by studying the relationship between airport connectivity and knowledge complexity. An approach that, to our knowledge, is the first to study this relationship. More generally, it gave insight into the determinants of regional knowledge complexity. Whereas previous research has established the principle of regional knowledge complexity and based on patent data regional knowledge complexity could even be accurately calculated, its determinants have mostly been neglected in previous research. With this study we started to investigate the determinants of regional knowledge complexity. However, only a limited amount of determinants concerning GDP, population and the share of high-tech employees have been investigated during this research. Whereas we already determined that the share employees in technology and knowledge-intensive sectors have a positive impact on the knowledge complexity of a region, several others socioeconomic factors should be considered. Socioeconomic factors, like higher education levels and unemployment rates, are especially relevant since these factors also have strong practical implications for, among others, the European Union.

5.2 Strengths and limitations

Whereas former research has merely focused on the concept of regional knowledge complexity, this research is one of the first to touch upon the determinants of regional knowledge complexity. However, the number of determinants studied have been limited and only airports centrality measures have been taken into consideration. The study indicated that airport degree, and to a lesser extent, betweenness might be sound indicators for the technological diversity of a region, but not for ubiquity. These results provide guidance to policy makers in terms of allocation of aviation funds, but also opens up a new field of study.

It is noteworthy that the analysis was performed on a small data sample and that the relations found were, if significant at all, only merely significant. Based on these results, the study should be repeated based on a more extensive data set to either verify or falsify our result and gain a more thorough understanding of the effect of airport connectivity on regional knowledge complexity.

5.3 Practical implications

This research provides a rationale to policy-makers regarding the allocation of financial resources to specific airports and aviation in general. Previous studies indicated that complex knowledge indeed contributes to economic growth (Hidalgo & Hausmann, 2009). The study of Hidalgo and Hausmann showed that the Index of Economic Complexity - a measure of the knowledge in a society as expressed in the products it makes - beats often used competitiveness measures by a tenfold in predicting economic growth for the following decade. These results underline the strong positive impact of knowledge complexity on economic growth and therefore the importance of knowledge complexity to regions or countries.

Based on the findings in this study, attempts to increase to connectivity of an airport will most likely also results in an increase of the technology diversity of a region, contributing to its knowledge complexity and in turn stimulating economic growth. Increasing the knowledge complexity of regions can strongly contribute to minimization of economic inequality between regions in Europe. According to our study, a viable way to do so would be by increasing the airport connectivity of regions that lack behind in terms of economic welfare.

In this study, we calculated the knowledge complexity of all European regions. This valuable data can assist the EU in targeting their aviation funds to specific regions that lag behind in terms of knowledge complexity and, based on the findings in our research, more specifically to regions that lag behind in technology diversity. We analyzed the 50 regions with the lowest technology diversity and found that a large share of those regions was located in eastern Europe, and more specifically in Romania, Slovakia and Czech Republic (a part of these regions are indicated in figure 2, highlighted in red). These findings are not surprising since these countries are mostly centered around agriculture, resulting in a low patent output of those regions. Hence, these regions have a low knowledge complexity. Another country that we identified to be lagging behind in terms of diversity is Portugal.

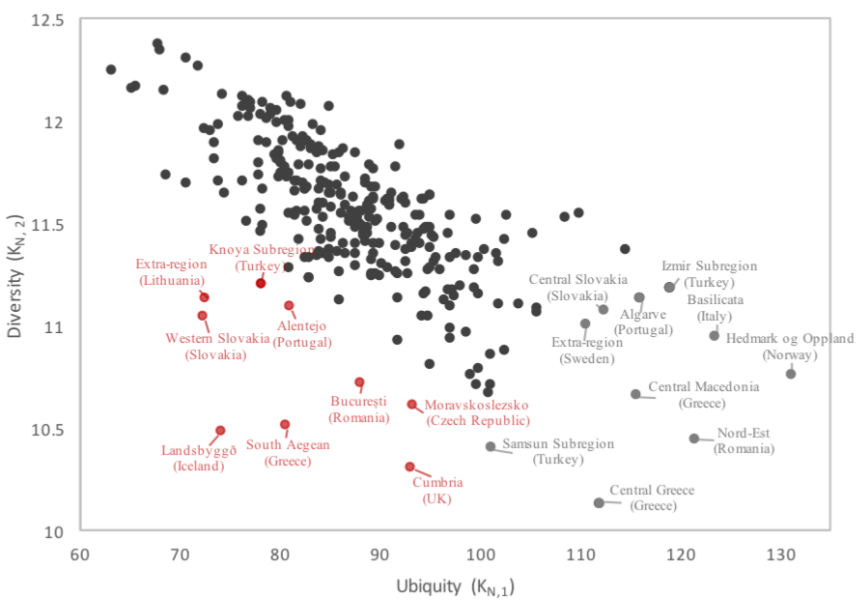


Figure 2. Technological diversity and ubiquity of NUT-2 regions. The red dots indicate the region with relatively low diversity, and the light grey dots indicate the region with relatively low technological diversity and high ubiquity.

Allocating financial resources to these specific countries or regions to lower both within- and between-country economic inequality is advisable based on the findings in this report. The European Union has two main aviation funding programs: the European Regional Development Fund (ERDF) and the Cohesion Fund (CF). Current aviation funds have mainly been allocated to Spain, Greece, Poland and Italy, collectively receiving 75 percent of the total amount of funding (European Court of Auditors, 2014). Whereas it makes sense to allocate funding to Greece and Poland based on the low KCI values - and patent diversity - of specific Polish and Greek regions, the large funds allocated to Italy and Spain could be allocated more efficiently.

The countries that we identified to be lagging behind in terms of technology diversity (Romania, Slovakia, Czech Republic and Portugal) together only receive a minor share of 6.1 percent of the total funding, where specifically the funds allocated to Slovakia (0,15%) and Romania (1,44%) are negligible (European Court of Auditors, 2014). Based on the findings in our report regarding the effect of airport connectivity and technology diversity, it would be advisable to reconsider the allocation of aviation funds in favor of these regions.

However, these recommendations might be conflicting with the climate related goals set out by the European Union. It is well-known that aerial travel or transport is one of the most climate-damaging ways of connecting two actors, and alternatives have to be considered. Another means of travel and transportation that comes to mind when one strives to travel far, fast and climate-friendly is the train. For this to be a real alternative, however, not only night trains but especially international connections instead of mostly merely national lines with a connecting train in the next country are required (Ottermann, 2014). One could thus put the research design and most of the data to further use for a similar research project conjecturing a similar link for train connectivity instead of airport connectivity. Should the result prove to be the same, one would have a result suggesting financial- and policy-oriented investments into the Intra-European railway system to not only increasing economic welfare but also countering climate change.

6 Conclusion

The following research question was addressed in this study: *To what extent does airport connectivity contribute to the development of regional knowledge complexity in Europe?* In order to answer this question, we first calculated a Knowledge Complexity Index (KCI) representing the knowledge complexity of a region. Then, flight data was analyzed to calculate the betweenness and centrality degree of airports within Europe. An OLS regression was performed to quantitatively study the effect of airport connectivity, indicated by the centrality measures betweenness and degree, on the regional knowledge complexity. The results indicated that there is no significant relationship between an airports degree and betweenness and the knowledge complexity of the corresponding region. However, once we split our knowledge complexity index into the sub-dimensions *ubiquity* and *diversity* we found that an airports degree and betweenness have a significant positive relation with a regions patent diversity. The effect of degree on knowledge complexity being larger than the effect of betweenness. The analysis was based on a small number of observations and increasing the number of observations by excluding a control variable made the effect of both airport centrality on knowledge complexity insignificant. Therefore, future research should be performed based on more extensive datasets in order to get a more thorough understanding of the effect of the connectivity of airports on regional knowledge complexity.

With this study we started to investigate the determinants of regional knowledge complexity, which have been mainly neglected in previous research. Knowing that airport connectivity is a determinant of regional knowledge complexity can provide a rationale for policy makers regarding the allocation of aviation funds. Based on our findings, we would advise to reconsider the allocation in favor of regions in Romania, Slovakia, Czech Republic and Portugal, which lack behind in terms of regional knowledge complexity. If allocated efficiently, these funds can increase the regional knowledge complexity of those regions, contributing to the decrease of both within- and between-country economic inequality.

7 References

Acs, Z., & Sanders, M. (2011). Patents, knowledge spillovers, and entrepreneurship. *Small Business Economics*, 39(4), 801-817. doi: 10.1007/s11187-011-9322-y

Allen, T. (1977). Managing the flow of technology. Technology transfer and the dissemination of technological information within the R & D organization. *Cambridge, Mass.: MIT Press*

Balland, P.-A., Boschma, R., Crespo, J. and Rigby, D. L. (2019). Smart specialization policy in the European Union: relatedness, knowledge complexity and regional diversification. *Regional Studies*, 53(9), 1252-1268. doi:10.1080/00343404.2018.1437900

Balland, P. A., & Rigby, D. (2017). The geography of complex knowledge. *Economic Geography*, 93(1), 1–23. doi:10.1080/ 00130095.2016.1205947

Baesu, V., Albulescu, C., Farkas, Z., & Drăghici, A. (2015). Determinants of the High-tech Sector Innovation Performance in the European Union: A Review. *Procedia Technology*, 19, 371-378. doi: 10.1016/j.protcy.2015.02.053

Bernela, B., Ferru, M., & Rallet, A. (2019). The impact of digital technologies on perceptions of proximity. Retrieved from: https://hal.archives-ouvertes.fr/hal-02051751/document

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

Brandes, U. (2001). A Faster Algorithm for Betweenness Centrality. *Journal of Mathematical Sociology*, 25(2), 163-177, doi:10.1080/0022250X.2001.9990249

Cantwell, J., & Salmon, J. (2018). *Chapter 5 The Effects of Global Connectivity on Knowledge Complexity in the Information Age. International Business in the Information and Digital Age, 123–137.*doi:10.1108/s1745-886220180000013006

Castaldi, C., Frenken, K., & Los, B. (2015). Related variety, unrelated variety and technological breakthroughs: an analysis of US state-level patenting. *Regional studies*, 49(5), 767-781. doi: 10.1080/00343404.2014.940305

Conventz, S., Derudder, B., Thierstein, A., Witlox, F. (2014). Hub Cities in the Knowledge Economy. London: Routledge.

Czarnitzki, D., Kraft, K., & Thorwarth, S. (2009). The knowledge production of 'R' and 'D'. *Economics Letters*, 105(1), 141-143. doi: 10.1016/j.econlet.2009.06.020

European Court of Auditors. (2014). EU-funded airport infrastructures: poor value for money. Retrieved from: https://www.eca.europa.eu/Lists/ECADocuments/SR14_21/QJAB14021ENC.pdf

European Patent Office (2014). FAQ - applying for a patent. Retrieved from: https://www.epo.org/service-support/faq/own-file.html

European Union (2017). Annual Analyses related to the EU Air Transport Market 2016. Retrieved from: https://ec.europa.eu/transport/sites/transport/files/2016_eu_air_transport_industry_analyses_report.pdf

Fleming, L., & Sorenson, O. (2001). Technology as a complex adaptive system: Evidence from patent data. *Research Policy*, 30(7), 1019–1039. doi:10.1016/S0048-7333(00)00135-9

Glaeser, E., Kallal, H., Scheinkman, J., & Shleifer, A. (1992). Growth in Cities. *Journal Of Political Economy*, 100(6), 1126-1152. doi: 10.1086/261856

Goh, K. I., Oh, E., Kahng, B., & Kim, D. (2003). Betweenness centrality correlation in social networks. *Physical Review E*, 67(1), doi: 10.1103/PhysRevE.67.017101

Grant, R. (1996). Toward a knowledge-based theory of the firm. *Strategic Management Journal*, 17(2), 109-122. doi: 10.1002/smj.4250171110

Green, R. (2007). Airports and Economic Development. *Real Estate Economics*, 35(1), 91-112. doi: 10.1111/j.1540-6229.2007.00183.x

Hausmann, R. Hidalgo, C., Bustos, S., Coscia, M., Chung, S., Jimenez, J., Simoes, A. & Yildirim, M. (2013). *The Atlas of Economic Complexity*. Cambridge, MA: MIT Press

Hidalgo, C. A., & Hausmann, R. (2009). The building blocks of economic complexity. *Proceedings of the National Academy of Sciences*, 106(26), 10570–10575. doi:10.1073/pnas.0900943106

Kararach G. (2014) Infrastructure and Connectivity. *Development Policy in Africa*. Palgrave Macmillan, London

Kline, M., & Boyd, R. (2010). Population size predicts technological complexity in Oceania. *Royal Society,* 277(1693), 2559-2564. doi: 10.1098/rspb.2010.0452

Kogut, B. & Zander, U. (1992). Knowledge of the firm, combinative capabilities, and the replication of technology. *Organization Science,* 3(3), 383-398. doi:10.1287/orsc.3.3.383

Laverde-Rojas, H. & Correa, J. C. (2019). Can scientific productivity impact the economic complexity of countries?. *Scientometrics*, 120(1), 267-282. doi: 10.1007/s11192-019-03118-8

Meghanathan, N. (2016). Assortativity Analysis of Real-World Network Graphs based on Centrality Metrics. *Computer and Information Science*, 9(3), 7-25. doi: 10.5539/cis.v9n3p7

Ottermann, P. (2014). End of the line for Europe's iconic night trains? Retrieved from: https://www.theguardian.com/world/2014/sep/12/europe-night-trains-sleeper-service

Pradhan, R. P., Arvin, M. B., Nair, M., Bennett, S. E., & Bahmani, S. (2019). Short-term and long-term dynamics of venture capital and economic growth in a digital economy: A study of European countries. *Technology in Society*, 57(1), 125-134. doi: 10.1016/j.techsoc.2018.11.002

Rocha, L. E. (2017). Dynamics of Air Transport Networks: A Review from a Complex Systems Perspective. *Chinese Journal of Aeronautics*, 30(2), 469-478. doi: 10.1016/j.cja.2016.12.029

Sapienza, H., Parhankangas, A., & Autio, E. (2004). Knowledge relatedness and post-spin-off growth. *Journal Of Business Venturing*, 19(6), 809-829. doi: 10.1016/j.jbusvent.2003.06.002

Scherngell T., Jansenberger E. (2006) Patents, Patent Citations and the Geography of Knowledge Spillovers in Europe. In: *Innovation, Networks, and Knowledge Spillovers*. Springer, Berlin, Heidelberg, 233-250

Shepherd, B. (2016). Infrastructure, trade facilitation, and network connectivity in sub-Saharan Africa. *Journal of African trade*, 3(1-2), 1-22. doi: 10.1016/j.joat.2017.05.001

Song, M., & Yeo, G. (2017). Analysis of the Air Transport Network Characteristics of Major Airports. *The Asian Journal Of Shipping And Logistics*, *33*(3), 117-125. doi: 10.1016/j.ajsl.2017.09.002

Storper, M. (1992) The Limits to Globalization: Technology Districts and International Trade. *Economic Geography*, 68(1). 60-93. doi: 10.2307/1144041

Trajtenberg, M. (1990). A penny for your quotes: patent citations and the value of innovations. *The Rand Journal of Economics*, 21(1), 172-187. doi: 10.2307/2555502

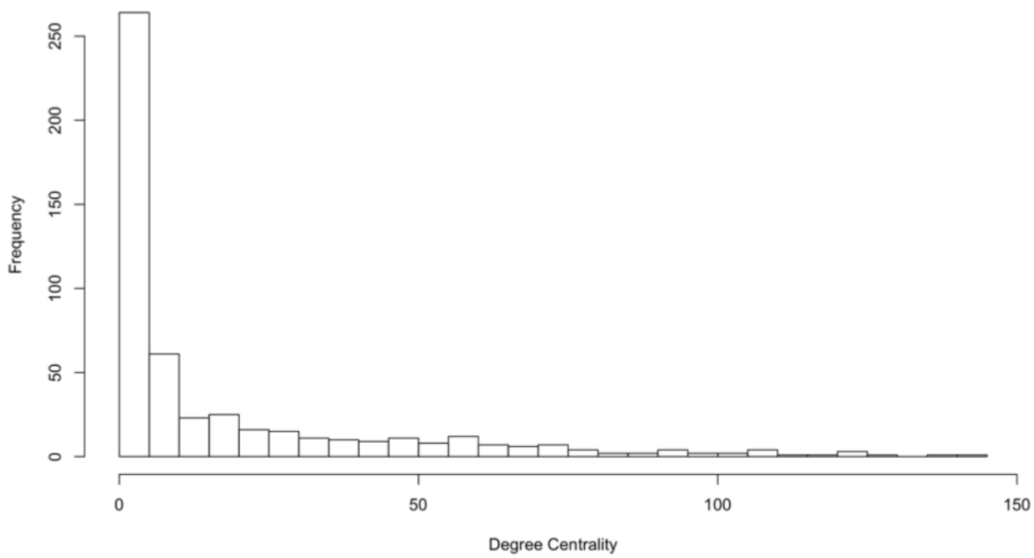
United States Patent and Trademark Office (USPTO). (2019). Data Visualization Center. Retrieved from: https://www.uspto.gov/dashboards/patents/main.dashxml

Veugelers, R. (2017). An innovation deficit behind Europe's overall productivity slowdown. Retrieved from: https://www.ecb.europa.eu/pub/pdf/other/ecb.ecbforumcentralbanking2017.en.pdf

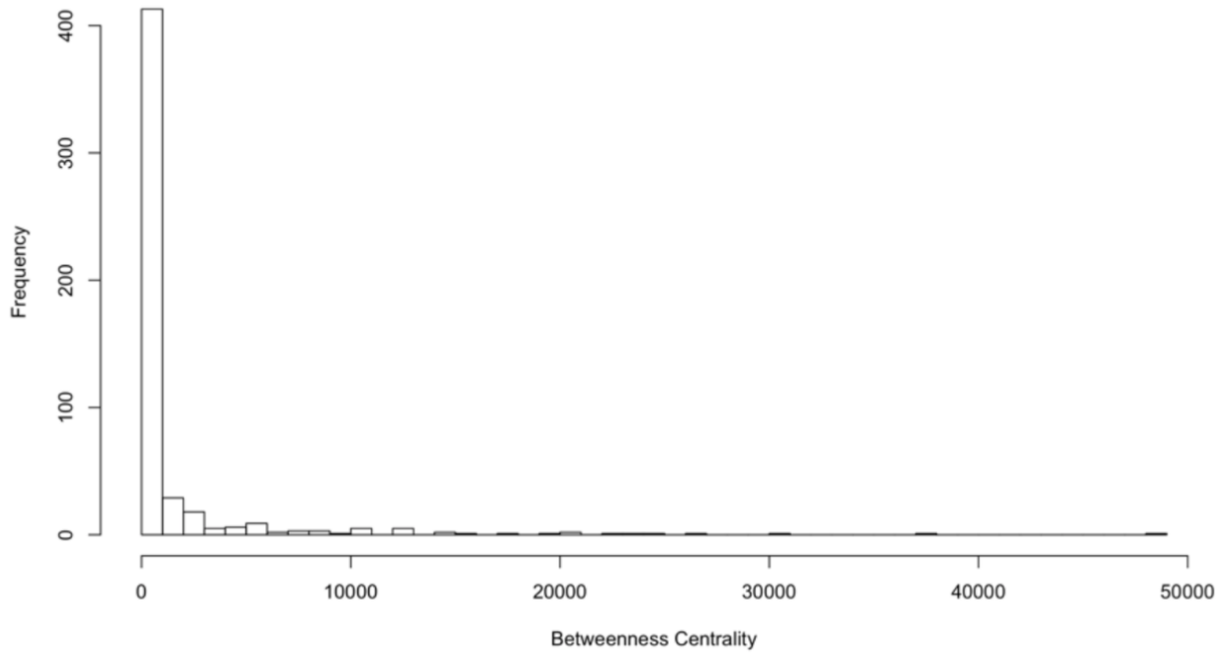
Van Zon, A., & Mupela, E. (2016). Connectivity and economic growth. *Macroeconomic Dynamics*, 20(8), 2148-2172. doi: 10.1017/s136510051500022x

Zhu, S., & Li, R. (2016). Economic complexity, human capital and economic growth: empirical research based on cross-country panel data. *Applied Economics*, 49(38), 3815–3828. doi:10.1080/00036846.2016.1270413

8 Appendix



S1: Value distribution of degree centrality



S2: Value distribution of betweenness centrality