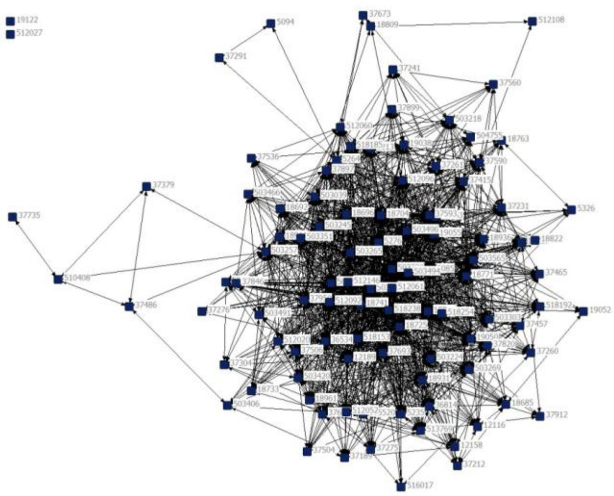
gatekeepers

in innovation networks



Who are the gatekeepers in innovation networks of the German biotechnology sector?

by

Laurens Hulshof

supervisors

Andrea Morrison

Tom Broekel

Master thesis Economic Geography - Faculty of Geosciences, Utrecht University

A B S T R A C T

In this study, the process of innovation through knowledge transfer is investigated by identifying gatekeepers in cognitive innovation networks in the biotech industry in Germany, using network analysis on records from an unprecedented combination of subsidy programs of the European Commission (FP6) and the federal German government. The results indicate that geographical distance between organizations does not act as a boundary for collaboration between the organizations that take part in these subsidy programs: Most actors are well connected beyond their own region. The gatekeepers are organizations that play an important role in connecting all organizations of the cognitive network. All gatekeepers are found to be organizations with a public mission (universities or research organizations), most gatekeepers are very active in the subsidy programs, taking part in many different collaborative projects, but the decisive characteristics of gatekeepers have yet to be discovered. Government and private organizations may utilize these conclusions in fine-tuning innovation (stimulation) policies and strategies.

index

1. introduction 5

1.1 Context of this research: innovation networks 5

1.2 Stimulating innovation: the role of gatekeepers 6

1.3 A network analysis: Gatekeepers in Biotechnology innovation networks, funded by European and German Subsidy Programs 7

1.4 Research questions 8

1.5 Contents of this research 8

2. theory 9

2.1 Context of gatekeeper research: ‘innovation networks’ 9

2.1.1. Innovation systems 10

2.1.2. Network analysis 10

2.1.3. Cognitive distance 12

2.2 Gatekeepers 14

2.2.1. Definition and characteristics in existing literature 14

2.2.2. Definition and characteristics for this study 17

2.3 Innovation in emerging markets and biotech 18

2.4 Policy implications 19

3. methodology 21

3.1 Quantitative approach of research question 1 and 2 21

3.1.1. Selecting the datasets 22

3.1.2. Assembling material for my theoretical framework 26

3.1.3. Preparing the datasets for analysis 27

3.1.4. Analyze the networks FP6 (1) and GER (2) separately 29

3.1.5. Identify the gatekeepers within each network 31

3.1.6. Combine the two networks through step 4 and 5 34

3.1.7. Compare the gatekeepers from the combined network (3) with the identified gatekeepers from network 1 and network 2 35

3.2 Qualitative approach of research question 3 37

3.3 Chapter summary 38

4. results 39

4.1 Step 4: analyze the networks FP6 and GER separately 39

4.2 Step 5: identify the gatekeepers within each network 45

4.3 Step 6: combine the two networks through step 4 and 5 48

4.4 Step 7: compare the gatekeepers from the combined network (3) with the identified gatekeepers from network 1 and network 2 49

4.5 Qualitative approach 58

4.6 Final remarks 61

5. conclusions 62

6. discussion 66

acknowledgements 69

literature 70

Image on front page: complete model of actors in the Sixth Framework Program selected for this study, visualized in Netdraw (Borgatti, 2002).

1. introduction

Are there specific conditions that stimulate innovation, and can they be purposefully created by means of policy? That is the question that has kept me fascinated during my study of economic geography. Figuring out how the innovation process works can be of great value, for the advancement of science and the strengthening economic systems, as can be seen by the wide variety of government policies aimed at the stimulation of innovation through subsidy programs. Governments (and their constituencies) can benefit from research into the innovation process by learning how to effectively enable innovators. Universities, research organizations and firms can apply new knowledge about how innovation works as well, in order to design their innovation strategies more effectively.

In this study, records of government issued subsidy programs are used to map networks of organizations that work together on innovations in the biotechnology industry. These collaborations may indicate which organizations play key roles in the innovation process. The gatekeepers are organizations that make knowledge flow possible through this network of collaborating organizations. By finding and analyzing gatekeepers, I hope to add to the understanding of the process of innovation by identifying its key players.

1.1. context of this research: innovation networks

There is a number of approaches to the research on innovation, most of which concern the processes of knowledge transfer. Innovation usually involves multiple organizations, that through collaboration share risks, access new markets and technologies and pool complementary skills (Powell et al., 1996). When organizations work together on innovations, sharing their skills or knowledge, a network of collaborating actors arises, through which knowledge is able to travel. Seen this way, an important part of the innovation process essentially takes place in networks of knowledge sharing people or organizations. By studying the way these networks are structured and how they evolve, comprehension of the process of innovation can be expanded and possibly influenced more effectively.

The way knowledge passes on between organizations has been studied by researchers from a variety of disciplines on different scale levels. In this study I make use of a combination of concepts and tools from several of these disciplines, in order to find some of the key players (gatekeepers) involved in knowledge transfer. Essentially, innovation systems are studied, a concept that includes economic as well as geographic constructs (Lundvall, 1985; Freeman, 1995). These innovation systems are studied using a tool that has its roots in sociology, called network analysis (Wellman, 1983). Using network analysis, the networks of innovation can be mapped by uncovering links between organization through which knowledge may transfer. The organizations in this knowledge network participate in subsidy programs that stimulate innovation, hence the word innovation network is used. Organizations that have a position of interest in the innovation network, and the characteristics of these organizations are the focus of this study.

1.2. stimulating innovation: the role of gatekeepers

Recent research (Giuliani & Bell, 2005; Morrisson, 2008; Graf, 2011; Morrison et al., 2012) on knowledge transfer focuses on the role of *gatekeepers* within collaborative networks. The gatekeeper organizations have a position in a network that is important for the overall connectivity of the network. They connect organizations that would otherwise not be connected, meaning that knowledge cannot pass between these actors without the gatekeeper.

Although it is not new (in social network analysis these gatekeepers are mostly referred to as brokers, for example in Haythornthwaite, 1996), the role of gatekeeper in a network is definitely underexposed in innovation network research (Ter Wal & Boschma, 2009). In a firm level research, Morrison (2008) investigates the role of gatekeepers as leaders of industrial districts, firms that act as a filter for local knowledge distribution. They are actors who, through their specific position in a network, have a key role in knowledge diffusion in that network.

On the firm level this can mean that an organization is the only one in its regional district that has connections to actors outside of that district. Fritsch (2008) found that public research organizations, like universities, usually have the role of gatekeeper in networks, due to their combination of intensive local and global connections and a public mission (Fritsch, 2008). In this study I attempt to make an addition to the toolbox of network analysis by further conceptualizing the role of the gatekeeper of knowledge on a structural network level. This means that a gatekeeper does not necessarily ‘keeps the gate’ of a territorially bounded unit, but has a position in a cognitive innovation network. This way, the role of geographic distance in knowledge transfer is assessed as well.

When gatekeepers exist in a network, they are important for the connectivity of the network as a whole. By getting to know more about the characteristics of gatekeepers in these networks, I hope to be able to add to the knowledge about innovation processes.

1.3. a network analysis: Gatekeepers in Biotechnology innovation networks, funded by European and German Subsidy Programs

The initial supervisors for my thesis, Andrea Morrison and Tom Broekel from Utrecht University, were able to provide me with a relatively unexplored set of network data on subsidies for collaboration of organizations in biotechnology, focused on Germany. The dataset of this study consists of these two subsidy programs, the EU Framework Programme 6, and a similar German subsidy program, which have been filtered for organizations in the biotechnology industry only. In order to receive funding from these programs, it is required for the organizations to collaborate with a certain number of other (international) organizations in the biotechnology sector. The projects in which the organizations collaborate provide the links between them, organizations that participate in more than one project connect all participants of all projects with each other. This way, a network is mapped.

The focus of this study is on German led subsidy projects in which organizations collaborated between 2003 and 2007. These restrictions are mainly the result of choosing the Sixth Framework Programme, which took place in that time period, and combining it with the German program, which includes only German organizations. Combining these programs presents a comprehensive view on the collaborations in the biotechnology industry in Germany, including collaborative links with non-German organizations.

In this study, the gatekeepers of both separate programs are identified, analyzed and compared. The data from the two programs is then combined, and the gatekeepers are again identified by the same characteristics, in order to verify the characteristics. The details about the EU Framework Programme 6, the German subsidy program and their comparison will be discussed in the methodology chapter.

1.4. research questions

The above has lead me to the following research questions:

1. *Which organizations can be identified as gatekeepers in the innovation networks of German led projects within the EU Framework Programme 6 and projects within the German subsidy program between 2003 and 2006 in the biotechnology sector? What are their characteristics? To what extent does the geographic position of these gatekeepers contribute to their role within the network?*

2. *If any, what are the similarities and differences between the characteristics of the gatekeepers in the EU Framework Programme 6, the German subsidy program and the combined network of both programs?*

3. *Are the results generally applicable in innovation networks of emerging technology markets and what are possible policy implications?*

1.5. contents of this research

In chapter 2, I explain the key concepts of my research questions in using existing literature, and I explain how this study relates to that research. I combine elements from the innovation systems theory with the concept of cognitive proximity, using network analysis as a tool to be able to answer the research questions. Then, existing literature on gatekeepers is discussed, in order to define gatekeepers for this study. Finally, some literature on the biotechnology industry and possible policy implications are discussed to provide a framework for the research subject.

In chapter 3, I expand on the method of research, that I utilize to answer the research questions. I will discuss the choice of using quantitative research measures for the first two questions, followed by an explanation of the seven steps that I will take to answer them. Then I explain the qualitative methods used in answering the third question.

In chapter 4, I will execute the research as described and explained in chapter 3. The gatekeepers in the networks at hand can be found and analyzed. In chapter 5, I assemble the analysis from chapter 4 in order of providing answers to the research questions.

In chapter 6, I critically assess the present thesis and provide an outlook for further research on the subjects discussed.

2. theory

introduction

In this chapter, I explain the key concepts of my research questions in using existing literature, and I explain how this study relates to that research.

To be able to explain the concept of gatekeepers I first elaborate in paragraph 2.1 on the context of their role in this research, namely innovation networks. Also, I denote the focus on geographic proximity in gatekeeper research so far, and why it is relevant from a theoretical perspective to focus in this study on cognitive proximity, before going into geographical proximity (research questions 1&2).

Then, in paragraph 2.2, I explain the definitions and characteristics of gatekeepers in existing research, and how this study relates to those definitions and characteristics (research question 2).

Finally, in paragraph 2.3 I discuss the significance of research on emerging markets in general, and the biotechnology industry specifically, according to existing literature in economic geography. I will explain how this study relates to that, to be able to give shape to policy implications (research question 3).

2.1. context of gatekeeper research: 'innovation networks'

Firms are not walled off like medieval kingdoms, but they are part of intricate networks of cooperation, even with competitors (Powell 1990). Sharing the risks of their research & development (R&D) activities is not the only reason for inter-organizational cooperation, firms also widen their scope to new markets and are able to make use of complementary skills of other firms and organizations (Hagedoorn, 1990; Graf, 2006). The inception of innovation is therefore often to be found between organizations, instead of within them (Powell, 1990; Graf, 2011). This makes innovation networks, and the organizations that are linked to each other through collaboration, interesting research subjects.

There is a number of approaches to the research on innovation, most of which concern the processes of knowledge transfer. The way knowledge passes on between organizations through collaboration has been studied by researchers from a variety of disciplines on different geographical and institutional scale levels (Ozman, 2009). In this research, I will combine network analysis with the study of innovation systems. Merging these analytic approaches, I refer to a network of organizations that may or may not collaborate with a purpose of innovation through collaborative R&D as *innovation networks*. Below, I explain the theory of the innovation systems approach, I discuss network analysis and I describe how both are applied in this study.

2.1.1. innovation systems approach

The innovation systems approach proposes successful innovation results from an effective collaboration between knowledge generating, knowledge transferring, and knowledge exploiting actors in a nation or a region (Kauffeld-Monz & Fritsch, 2008). Freeman (1995), via List (1840s) and Lundvall (1980s, 1992), considered national and regional innovation systems to be the drivers of economic development and growth. Innovation systems consist of technical and scientific institutions, the education system, government policy and culture in a country, that form a fundamental knowledge network that drives innovation (Freeman, 1995). Analytical studies on innovation must encompass the whole system, meaning that every entity (actor or institution) that plays a role in the innovative performance of a system should be included in analysis (Lundvall, 1992; Edquist, 1997).

Mainly because of the advantages of spatial proximity between organizations, the benefits of *regional* innovation networks (e.g. Braczyk & Cooke, 1998; Cooke, 2001) are considered among the main causes of knowledge transfer through localized knowledge spillovers, which form an important explanation of spatial clustering of innovative activity (e.g. Krugman, 1991; Audretsch & Feldman, 1996; Breschi & Lissoni, 2001; Malmberg & Maskell, 2002, and see text box 2.1).

A variety of economical and geographic concepts can be utilized with innovation systems functioning as a conceptual vehicle in which knowledge is generated and diffused. Among others, concepts of (localized) knowledge spillover (e.g. Jaffe et al. 1993; Audretsch & Feldman, 1996; Breschi & Lissoni, 2001), evolutionary concepts in innovation systems (e.g. Lambooy & Boschma, 2001; Gilsing & Nooteboom, 2006; Glückler, 2007), the role of (related) variety (e.g. Boschma & Iammarino, 2007; Frenken et al., 2007), and complexity in innovation systems (e.g. Katz, 2006) have been explored within the framework of the innovation system. In this research I focus on the notion of the innovation system as a network of organizations, in which some organizations have the role of gatekeepers.

2.1.2. network analysis

The process of innovation usually involves a large number of actors, organizations that may or may not cooperate with each other (Ozman, 2009). By focusing not on the singular actors (firms, research institutes, universities), but on the links through which knowledge flows between them, a network can be exposed of actors that play parts in the innovation process (Freeman, 1991). Network analysis uses quantitative techniques to describe and analyze the ties between nodes (actors) that are linked in a network. These techniques are established in graph theory and statistics, making a structural analysis of networks possible (Wasserman & Faust, 1994).

The tool of network analysis was originally applied by sociologists, ethnologists and anthropologists to describe and analyze personal interactions (Wellman, 1983; Maggioni, 2009) since the 1960s, but has grown in popularity in recent years, when advancements in computer software allowed very large and complex data sets to be empirically assessed (Ter Wal & Boschma, 2009). During the past decade and a half, a growing number of economists and geographers became convinced that network analysis allows for an appropriate conceptualization and evaluation of complex relations and knowledge flows between agents, industries, regions and countries (Grabher & Ibert, 2006; Maggioni, 2009; Ozman, 2009). Network analysis makes empirical assessment of knowledge flows possible, to be studied in broad strokes as well as in detail, on different scale levels and through time, making it a dynamic research tool (Ter Wal & Boschma, 2009).

Networks can be analyzed on different levels, namely on a dyad level, a node level and a structural level. Research on a dyad level focuses on pairs of actors, for example the chance that two organizations are connected when they have a similar knowledge base showing the importance of cognitive distance in that network (Broekel & Hartog, 2011). The node level refers to characteristics of the units of analysis, being the actors in the network. The size of the organization for example, may play a large role in the number of ties it is likely to have in a network (Boschma & Ter Wal, 2007). The structural level entails the entire network, on its characteristics as a complete network of actors. On this level, characteristics can be attributed to nodes in the network simply because of their position in that network. The actors that hold these positions of interest may then be analyzed separately on the node and/or dyad level, while they already have been singled out because of their structural network characteristics. Analytical emphasis is put on the structural effects of an actor's position in the network (Glückler, 2007).

This research will focus primarily on the structural network level, for two reasons. Firstly, analysis of innovation systems must encompass the whole system (see 2.1.1). The empirical tool of network analysis allows analysis on a structural level, even in large scale networks, which this study is. Secondly, at the structural level gatekeepers can be located using a purely technical definition in a network with cognitive links (see 2.1.3). Then, the gatekeepers that are found are analyzed on the node level. This way, applying an ‘outside-in’ method, any geographical similarities as well as other corresponding characteristics between gatekeepers can be found separately without any previous assumptions.

2.1.3. cognitive proximity approach

In economic geography, the effect of physical proximity is the most widely used concept for evaluating network formation (Glückler, 2007). But the proximity between organizations can have the form of several difference dimensions, geographical distance being one. However, the importance of geographical distance has been questioned by several researchers (e.g. Breschi & Lissoni, 2001; Boschma, 2005; Ponds et al. 2007).

According to Boschma (2005), there are five forms of proximity with which the distance between organizations can be indicated. They are cognitive, organizational, social, institutional and geographical proximity. These proximities may play a part in the quality of interaction between organizations (Boschma, 2005). In this research gatekeepers will be designated in a network of cognitive relations, which will be compared to their geographical characteristics, making cognitive and geographical proximity the focus of research. Cognitive proximity signifies the extent to which two organizations share the same knowledge base, and geographical proximity indicates the physical distance between two organizations (Frenken & Boschma, 2009).

Innovation networks have been shown to be structured unevenly, meaning that some nodes in the network are not as highly connected as others (Ozman, 2009). Geographical proximity between organizations affects, but does not fully explain this uneven distribution of network links (Breschi & Lissoni, 2003; Glückler, 2007). Knowledge networks are non-territorial: Knowledge transfer takes place between local actors, but may also span across the globe (Ter Wal & Boschma, 2009). Explaining the process of innovation needs a non-territorial vehicle, categorizing actors on the basis of their structural relationships with each other (Wellman, 1983). Among others, Nooteboom et al. (2007) have demonstrated that cognitive proximity can be seen as an important determinant in R&D alliances.

At the dyad level of network research, a common practice is collecting and comparing knowledge bases of different organizations, and comparing these to the links in a network to test the importance of the knowledge base to tie creation between actors in that network. Much of these studies require primary data sources (e.g. Giuliani & Bell, 2005; Morrison, 2008; Boschma & Ter Wal, 2007), like surveys or interviews in which respondents indicate the details on their connections with other actors, which prove impractical in large scale networks analysis that encompasses the whole system (Ter Wal & Boschma, 2009).

Because this is a large scale network study, the structural level is an appropriate starting point. The collective knowledge base is the starting point for the links in this network: The organizations in this network are linked through collaborative subsidy programs. The links are the projects in which the actors collaborate, which I assume as a collective knowledge base, i.e. cognitive proximity. In this study cognitive proximity is assumed between organizations that collaborate in R&D projects in subsidized government innovation programs. Not only is there a certain and non-territorial proximity between collaborating organizations, by contract all results from the projects may be unlimitedly used by all the participants of the projects. This means that their collective knowledge base is the same at least on the project in question. Network analysis is utilized to map the network that results from these collaborative R&D projects.

Cluster theory and innovation

In this study I focus on innovation systems, which exist alongside of cluster theory in the economic geographical paradigm. As explained in paragraph 2.1.2, I do not place my research in a geographical framework *a priori*. However, it is important to note that most gatekeeper literature is grounded within cluster theory. Although I do not apply cluster theory as a basic framework, instead I approach innovation networks from a cognitive proximity perspective, this study certainly refers to cluster theory on several occasions.

Tacit, or un-codified knowledge is suggested to be able to travel limited distances, making geographic distance a barrier for knowledge flows (Krugman, 1991; Von Hippel, 1994). In cluster theory, spatial proximity itself is argued to be the driver for innovation and economic development. Tacit knowledge is shared easily through *local buzz* (the effective circulation of knowledge, facilitated by social and cultural homogeneity (Bathelt et al., 2004; Morrison et al., 2011), face-to-face contacts, shared institutions and labor mobility (Jaffe et al., 1993; Feldman, 1994; Audretsch & Feldman,1996; Breschi & Lissoni, 2003; Ter Wal & Boschma, 2009).

However, firms within a cluster can only learn a limited amount from each other (Nelson & Winter, 1982), so there is the risk of *lock-in,* meaning that a deficiency of knowledge flows from outside the cluster leads to substandard innovative practices (Grabher, 1993; Uzzi, 1996; Boschma, 2005). Therefore, inter-regional knowledge flows (*Global pipelines* see also 2.2.1) are advantageous for regional innovation processes (Camagni, 1991)

In the next paragraph, I will discuss definitions and characteristics of gatekeepers which stem from cluster theory studies. From this theory on gatekeepers, which is set in a geographical framework, I draw parallels to the role of gatekeepers framed solely in cognitive proximity.

Text box 2.1: cluster theory

2.2. gatekeepers

In this paragraph I present and adopt previous research that has been conducted on gatekeepers. In paragraph 2.2.1. I chronically discuss definitions as well as found conclusions and characteristics of gatekeepers in those studies. Based on this literature I will define gatekeepers and their characteristics for this research in paragraph 2.2.2.

2.2.1. definition and characteristics in existing literature

In network analysis, gatekeepers are nodes that form a connection between disorganized others (Haythornethwaite, 1996), or more specifically: actors in a network that hold a position between other actors that are not linked directly (Nooteboom, 2003). Actors that hold this position in a network may be ascribed a certain level of power, because they are in the positions to extend or withhold certain knowledge or information from other actors (Wellman, 1983).

The concept of 'technological gatekeepers' of knowledge as actors that combine the advantages of external and internal connections was introduced by Allen (1977). Therefore, his conception serves as a starting point for discussing literature about gatekeepers in networks. Allen was concerned with the way an R&D organization could keep up with technological development in their industry or sector. One of the ways to do this is to encourage members or employees of the organizations to arrange direct formal or informal contact with outsiders who possess desired technological information (Allen, 1977). Persons who had a high degree of long term informal contact with colleagues outside of their organization were labeled technological gatekeepers, and their characteristics were described:

- They were 'high communicators';
- Who were highly read up on the professional literature.;
- They were also high technical performers;
- And were mostly represented at the lower hierarchical levels of the organization and not officially recognized in the role of gatekeeper.

Overall, they could be described as having excellent knowledge of external developments, which they could apply to their own organization or translate to internal co-workers. The analogy of the firm was translated to different units of analysis for later studies.

A jump in time in gatekeeper literature leads to Giuliani & Bell (2005). In their methodologically extensive research, network theory is adapted for economic geography, by empirically mapping knowledge networks in a district like context (Morrison, 2008): a Chilean wine cluster. The study has the explicit goal of finding gatekeepers in a cluster, by conceptualizing innovation process within the cluster, which was more of a black box before the appliance of network analysis (Giuliani & Bell, 2005).

Following the definition of Giuliani & Bell (2005), the technological gatekeeper is a firm with a locally central network position in terms of knowledge transfer, that is also strongly connected to external sources of knowledge (Giuliani & Bell, 2005). The gatekeeper monitors the external environment and translates technical information into practical knowledge to the local organizations. The gatekeeper operates on the border of a cluster, preventing lock-in by gaining, translating and dispersing knowledge from outside connections to organizations within the cluster, much like the employee in Allen (1977) did.

Seen this way, the gatekeeper in cluster literature operates between the space of places (being in the right place) and the space of flows (being part of the right network), as suggested by Castells (1996). It is the link between global pipelines and local buzz (Bathelt et al., 2004), in which global pipelines are (often not spatially bounded) connections and local buzz is the the effective local circulation of knowledge, facilitated by social and cultural homogeneity (Bathelt et al., 2004; Morrison et al., 2011). This means the gatekeeper can absorb knowledge and transform it into production activities. The gatekeeper firm has a workforce that is skilled in knowledge absorption, and has a structure that is able to channel knowledge within the firm, but also between firms (Giuliani & Bell, 2005).

In their research, the absorptive capacity and ability of transformation of knowledge of firms has been measured on the node level, by combining employee and company statistics on education and experience, among others, for every firm in the cluster (Giuliani & Bell, 2005). The position in the network has been investigated with network analysis, which is discussed in chapter 3 of this study. The firms in the Chilean wine cluster were shown to vary a great deal in terms of their connectedness to the network, some firms were isolated, while others were very well connected.

In a study that uses roughly the same gatekeeper definition and characteristics, Morrison (2008) showed that a number of relatively large firms in a furniture district in the south of Italy were acting as gatekeepers for their cluster, which gave those firms a leading role in that district (Morrison, 2008). In his study, Morrison (2008) also argues that an actor is not a gatekeeper without the '*willingness to share'*, knowledge with local organizations. A gatekeepers' attitude depends on reciprocity with other members of the cluster (Morrison, 2008).

In a study by Kauffeld-Monz & Fritsch (2008), a methodology similar to that of the present study is used. Also like this research, the subject of analysis is the biotechnological sector. Gatekeepers are identified on the structural level, on the basis of network characteristics, but are applied in regional innovation systems in the biotechnology sector, wherein the gatekeeper has the function of preventing lock-in. Kauffeld-Monz & Fritsch (2008) use a similar definition as Giuliani & Bell (2005), with the gatekeeper as a 'boundary spanner' that injects globally acquired knowledge into the regional innovation process it is part of. The gatekeeper thus provides the regional system with a global outlook (Kauffeld-Monz & Fritsch, 2008). These gatekeepers have high quality links to global knowledge sources, as well as to local organizations, and willingness to share its acquired knowledge regionally. The study suggests that universities and, to a lesser extent, larger public research organizations often assume the role of gatekeepers, which may be of benefit to lagging regions because of their public orientation and international relations (Kauffeld-Monz & Fritsch, 2008).

A network research by Graf (2011) focuses on clusters in the biotechnology industry in four German regions. In similar fashion as Kauffeld-Monz & Fritsch (2008), these regions are compared and the characteristics of the gatekeepers are studied. Graf (2011) finds that large multinational firms can be of importance in clusters because of their extensive absorptive capacity, but that the lack of a public mission often nullifies the willingness to share. The size of an organization therefore does not matter as much as the presence of a public mission. Public research organizations have a public goal, but often are internally oriented and therefore have suboptimal absorptive capacity, making them less effective in external knowledge acquisition (Graf, 2011).

Some studies designate organizations that are not willing to share as 'external stars'. These are organizations that have strong external connections but limited links with cluster internal organizations due to, for example, the lack of incentive to locally share knowledge (e.g. Giuliani & Bell, 2005; Morrison et al., 2012).

The research of Morrison, Rabelotti & Zirulia (2012) focuses more on the internal cluster conditions. Characteristics of gatekeepers are that they are present in clusters with a mix of internal density and external linkages, and they connect social and cultural homogeneous networks. Global pipelines are useful to a cluster when there is a high quality local buzz, meaning an effective knowledge diffusion mechanism within a cluster, and the cluster is weakly endowed in terms of knowledge, meaning that there are no internal substitutes for the global pipelines (Morrison et al., 2012).

Additionally, it should be noted that gatekeepers could be defined differently in different research functionalities. A study focused on evaluating the Sixth Framework Programme (Malerba et al., 2006), uses the terms gatekeeper and 'hub' in a network, to differentiate between the knowledge brokers in one network (hub), and knowledge brokers between different networks (gatekeeper). In Malerba et al. (2006) the networks are European and global subsidy network much like the ones analyzed in the present study, and the focus is on 'gatekeepers' that connect the two scale levels. In the case of Malerba et al. (2006), the differentiation between the terms is functional to the research proposition. This study also utilizes a definition that is functional to the research, it is explained in paragraph 2.2.2.

2.2.2. definition of gatekeepers for this study

In this study, the characteristics and definition for gatekeepers are inspired by the definitions and characteristics discussed above, but also deviates because the unit of analysis is not a geographical entity. The discussed literature, in short, defines a gatekeeper as 'keeping the gate' of a spatial area. In this study, a gatekeeper is *a priori* a cognitive entity. Whether this position also proves to have a geographical significance will be investigated separately, for reasons explained. Network analysis inherently is a technical process. Therefore, the definition cannot be solely defined based on literature. Part of the definition is a methodological construct, which will be established in detail in the next chapter, in paragraph 3.1.5.

In this study, the definition of a gatekeeper in innovation networks is:

> An actor that holds a position in a network through which significant parts of the rest of the network are connected, and is willing to share knowledge with other actors. Removal of this actor results in significant reduction of knowledge diffusion in the network, giving the gatekeeper a potential position of power.

Text box 2.2: definition of a gatekeeper

This actor 'keeps the gate' between large parts of its network. The more connections between other actors it makes possible, that would not be connected without it, the more important that actor is to the cohesiveness of the network. In the next chapter, a network measure will be explained that calculates for every actor in the network how important it is corresponding to this definition. This measure is called 'flow betweenness'. 'Pure' gatekeepers are actors that are the only one that connect the other parts. In large networks these actors do not often exist, because there is almost always another way two actors are connected, be it via a longer route through the network. In this study, actors with a very high level of flow betweenness are called gatekeepers, even if there are more paths between two other actors. This will be expanded upon in the methodology chapter.

According to the studies in paragraph 2.2.1, in combination to the difference in definition stated above, characteristics of the gatekeeper in this study are:

- It holds an important position in the network with respect to other actors, concerning the knowledge diffusion in the network
- It has high quality links with both cognitively close actors in the network, as well as distant ones.
- It has high absorptive capacity and ability to translate tacit into codified knowledge and *vice versa*.
- It is willing to share knowledge between cognitively close and distant actors, for example because it has a public mission.

These characteristics will be the basis for the methodological concept of gatekeepers in this study. I will elaborate on this in chapter 3. In chapter 4, I will locate gatekeepers in innovation networks based on these characteristics.

2.3. innovation in emerging markets and biotechnology

In this paragraph, I discuss the history and significance of the biotechnology sector as a unit of research according to existing literature in economic geography, in order to offer a framework for this research. I will then describe the relation of this study to the relevant literature on biotechnology. The biotechnology sector has some characteristics which may be of relevance in answering research question 3.

The biotechnology industry can be described as a collection of technologies that has become influential in existing industries like agriculture, food-processing and human health (Pisano, 2002), in which recent advancements in molecular and cellular biology are applied (Powell et al. 1996). The industry has its history in genetic engineering in food processing and pharmaceutical industries, and is a science-based and knowledge intensive sector (Ter Wal, 2010).

The industry life cycle (Audretsch & Feldman, 1996; Nooteboom, 2000) of the biotech sector deviates from the cycle of typical traditional industries (Ginsling & Nooteboom 2006). In more traditional sectors, a period of radical innovation and codification of knowledge is typically followed by a focus on efficiency of production and distribution in the industry (Gilsing & Nooteboom, 2006), resulting only in minor further innovations, called *incremental* innovation (Callon, 2002). In the biotechnology sector, however, small firms seem to keep radically innovating. These firms are often university spin offs, which places them close to academic research while they have their focus on commercial potential (Lehrer, 2005). As it is maturing in the industry life cycle (e.g. Ter Wal, 2010), the biotech sector becomes increasingly codified, which increases the chance of knowledge spillovers (Gilsing & Nooteboom, 2006). A more detailed account of this process can be found in Gilsing & Nooteboom (2006) and Ter Wal (2010).

The biotechnology industry is a popular object for researchers in a variety of disciplines. Economic geographers study biotechnological innovation networks because of the high frequency of collaborations between organizations, which occurs because of the financially risky and time consuming process of innovation that is typical for knowledge-intensive and science based industries (Ter Wal, 2010). These collaborations can be studied economically, geographically and using network analysis, and can be useful for governments who wish to stimulate biotechnological activities. This will be discussed In paragraph 2.4.

Biotech activities have been found out to spatially agglomerate in a limited number of regions, forming clusters (Glückler, 2007; Kauffeld Monz & Fritsch, 2008; Wink, 2008; Ter Wal, 2010; Krafft, 2011). Fornahl, Broekel and Boschma (2011) distinguish seven biotechnology clusters in Germany, which will be used as reference in this study. The clusters are in: Berlin, Göttingen, Hamburg, Munich, Rhine-Main, the Rhineland and Rhine-Neckar. The results of this study will be compared to these biotech clusters in order to evaluate possible geographical patterns.

2.4. policy implications

In this research, I will use network analysis to study innovation systems. I examine cognitive relations between organizations by investigating networks of organizations in collaborative subsidy programs on a structural level. I will locate, describe and analyze gatekeepers in these networks. This way I contribute to the understanding of the role of the knowledge gatekeeper in innovation networks. If possible, I will draw policy implication from this analysis.

According to Porter (1998) nations and regions are competitive, and competitive advantage can be stimulated. The role of governments herein, is that of a catalyst and a challenger. Only firms can create competitive industries, governments can only provide the right circumstances (Porter, 1998, p. 200 – 204).

Learning about knowledge gatekeepers in innovation networks can be useful to firm and government policies, because innovation can play a significant role in the (international) competitiveness of firms, regions, countries and economic unions, by strengthening the scientific and technological knowledge base of its industries (Porter, 1998; Malerba et al., 2006). In this paragraph I briefly discuss the way this study may influence policy.

According to these studies, gatekeepers do not only hold a powerful position in networks, they also have certain characteristics that might be conducive for the connectivity of the network. These can be found in paragraph 2.2. Malerba et al. (2006) argue that hubs, which are similar to my definition of a gatekeeper, can be important in strengthening the scientific and technological knowledge base of industries. Existing hubs (gatekeepers in this study) should be encouraged, and organizations that already show some gatekeeper characteristics should be stimulated in the parts that they lack from being gatekeepers (Malerba et al., 2006). Morrison & Rabelotti (2012), also plead for the encouragement of global pipelines and a high quality local buzz through institutional and infrastructural support, because these cannot be taken for granted. Also, they argue that globally connected actors might need incentives in order to convince them to play the part of gatekeeper (Morrison & Rabelotti, 2012). This way, governments can be of assistance in making their region, country or union more competitive. In this study I will contribute to the study of gatekeepers to assist in learning about innovation networks. Understanding gatekeepers will hopefully aid successful stimulation of knowledge distribution by governments and organizations.

3. methodology

introduction

In this chapter I expand on the method of research, that I utilize to answer the research questions. I explain the entire process of conducting this research. To be able to answer research questions 1 and 2, I will use quantitative research methods. In paragraph 3.1 I will discuss this choice, followed by an explanation of the seven steps that I will take to answer these questions. Paragraph 3.2 will explain the method of the third research question, which has a qualitative nature. Finally, the chapter is summarized in paragraph 3.3.

3.1 quantitative approach of research questions 1 and 2

In order to answer research question 1 and 2 I use the quantitative methods of network analysis. Networks, being fitting conceptualizations of inter-organizational interaction and knowledge flows, can be empirically assessed using network analysis (Ter Wal & Boschma, 2009). Network analysis allows for the study of networks as a whole, even if they contain large amounts of actors.

In this study, a large and widespread group of interlinked actors in innovation networks is analyzed in order to find gatekeepers and learn about them. The actors in this study are all organizations that are active in a biotechnology collaboration project. They will be referred to as actors, nodes (when related to network analysis) or organizations. The data for the three innovation networks that I will analyze in this study are extracted from the public databases on two partly overlapping collaborative subsidy programs. The first two networks are created based on the two subsidy programs, the third is a network of the two subsidy programs combined into one large scale network. I will locate the gatekeepers among the actors in these three networks using technical network characteristics. The quantitative techniques of network analysis allows me to perform accurate analysis on this large scale. Because all the organizations that fit the requirements for this study are included, the study could be representative of a large population, while a detailed analysis of (actors in) the network is also possible.

I will follow the order of these steps in order to answer these two questions. Each of the steps is explained in a separate paragraph below.

<u>step 1:</u> Selecting the datasets (par 3.1.1)

<u>step 2:</u> Assembling material for my theoretical framework (par 3.1.2)

<u>step 3:</u> Preparing the datasets for analysis (par 3.1.3)

step 4: Analyze the networks FP6 (1) and GER (2) separately (par 3.1.4)
step 5: Identify the gatekeepers within each network (par 3.1.5)
step 6: Combine the two networks through step 4 and 5 (par 3.1.6)
step 7: Compare the gatekeepers from the combined network (3) with the identified gatekeepers from network 1 and network 2 (par 3.1.7)

3.1.1. selecting the datasets

The goal of this study is to find gatekeepers in a network of knowledge flows. Finding these actors can be achieved through social network analysis, because the gatekeeper by definition holds a certain position in a network. Social network analysis is focused on linkages between objects of study, instead of on the objects themselves. Instead of actors and their attributes, the focus is on actors and the relations between them (Hanneman & Riddle, 2005). By studying these linkages, researchers can study attributes of complete groups of actors that are demonstrably linked to each other. Another feature of social network analysis is that groups of subjects can easily be studied progressively through time, making it a dynamic technique of research (Glückler, 2007).

Social network analysis can be applied using primary and secondary data. Primary data is collected from the source, usually using surveys or interviews. This is considered the more statistically robust technique, because it is based on first hand evidence of linkages between actors. There are however some practical shortcomings, among which are limitations in scale and dynamism (Ter Wal & Boschma, 2009). In this study secondary data is collected, prompting different analysis techniques. Knowledge flows collected from records that imply linkages (secondary) can offer large quantities of data, but actual knowledge spillover is harder to prove.

Primary sources provide a snapshot of the current network, but records of secondary data can be traced back in time, making an ex post evaluation of networks possible, as well as the evolutionary (longitudinal) study of network dynamics over time (Ter Wal & Boschma, 2009, Frenken & Boschma, 2010). For studies with a longitudinal, or evolutionary focus see Glückler (2007) and Boschma & Frenken (2007), for a comprehensive discussion, see Ter Wal & Boschma, 2009 and Boschma & Frenken, 2010.

In this research I focus on one time period, making it a static study. The period in which this study is limited is the running time of the Sixth Framework Programme, which was between 2003 and 2007. I use this Framework Programme because it is the most recent one that is completed. Finding and analyzing the gatekeepers for the combined collaborative subsidy networks in one time period is a first step in this method of learning about the role of gatekeepers in innovation networks. Taking a static point of view may result in some biased outcomes, as the network positions of actors are likely to differentiate through time. Therefore, an evolutionary study through more time periods would be useful. It is possible to duplicate this study with previous and subsequent network data of the same organizations in these subsidy programs for the study of evolutionary, longitudinal network effects. Other installments of the Framework Programme would be a perfect foundation for this purpose. This study, however, retains a static focus on the period of one Framework Programme in order to determine if this exact type of research can add to the study of innovation networks and the role of gatekeepers herein. An evolutionary study would be too comprehensive for this purpose, as processing the data is a time consuming enterprise. If this type of study proves to be fruitful, a longitudinal study could be considered.

Data on patents and patent citations are a common source for this type of network research (e.g. Jaffe et al., 1993; Stuart, 1998; Balconi et al., 2004; Ejermo & Karlsson, 2004; Hoekman et al., 2009; Graf 2011). The secondary data source of patent records, however, include only actors (persons or organizations) that were involved in innovations that were deemed worthy to protect. While patent information contains a treasure of details about the inventors, collaborators, firms and citations to other patents, as well as the dates and sector of the invention (Ter Wal & Boschma, 2009), it only provides that information on innovations that were successful, in need of legal protection and financially worth a patent fee. In short, a considerable amount of information may be lost in using patent data. Additionally, network analysis is only possible with those patents that are linked to more than one inventor or organization.

In this research, I will uncover links between organizations by using subsidy records of collaboration programs. The diffusion of knowledge does not only flow though successful innovation. Different from patent data, collaborative subsidy programs contain links between actors that have worked together in projects, whether they have achieved innovation or not. Thus, this network of collaborating entities contains a category of actors that is broader in scope than a network based on patent records, while they still are part of a network in which collaboration is similarly intensive. The organizations that are present in this study have successfully applied for government grants, however, which means that the selected actors are definitely part of a collaboration project. Isolated innovation organizations or organizations that were unsuccessful in acquiring government grants are not part of this research.

These subsidies for research projects are issued by governments that require multiple organizations to collaborate in projects. A network can be uncovered with these records, collecting the linkages between the collaborating partner-organizations. These collaborations provide links between organizations that have undoubtedly worked together, thus making the subsidy records an appropriate source for network analysis.

Links for this study are the organizations that are listed together in projects that received funding. The participants of a project group work together on the project that is subsidized, so connections that enable knowledge flows between these organizations are present. In this study, the connections between the collaborating organizations in these project groups are viewed as possible knowledge flows. When organizations are involved in more than one project, they provide a link between all the members of all the projects they are in. This creates a possible knowledge network between every connected participant of the subsidy program.

Using secondary data, a very large network can be mapped out, by assembling all the participants of a subsidy program in one dataset. Unfortunately, the secondary nature of the data also makes it impossible to prove the existence of actual knowledge flows. Therefore, this study maps out a network of *possible* knowledge flows. Organizations that are collaborating in subsidy projects with the goal of innovating can be assumed to work together and share knowledge. The process and possible outcomes of the project are to be fully utilized by all participants of the project group, but it is possible that other knowledge is shared as well. It requires the addition of primary data, like surveys or interviews to prove these actual knowledge flows, as it is also possible that participants of a project do not interact at all. To collect primary data on collaboration for the entire network would prove to be too time consuming. It is possible to collect primary data on the organizations of interest, the gatekeepers in this study, for example. This however, is beyond the goals of the present research. In chapter 6 (the discussion) I will discuss this subject more thoroughly.

Organizations that are in a project together are linked directly to each other. Organizations that participate in more than one project form links between these project groups. From a network perspective, organizations as well as whole projects can be considered as nodes in the network, which provides multiple options for analysis. I will discuss this in detail in step 4, paragraph 3.1.4.

Linkages found in two different collaborative subsidy programs are used to answer the first two research questions: The European Commission Framework Programme 6 (FP6 in short), and the German federal government subsidy program (from now on also called GER). These programs are similar but autonomous subsidy programs, both intended to stimulate innovation in the high tech sector by awarding funding to joint research projects between different organizations.

‘FP6’

The framework programmes are the main policy instrument of the Community for research and technological development (RTD). The treaty of the European Union through the European commission intends to build a more competitive technological and scientific base for the industry in the European countries, by encouraging international research collaboration (Cordis website, 04-2012). Since the 1980s, successive Framework Programmes have tried to achieve these goals by supporting collaborative research (Malerba et al, 2006). The sixth FP outing is the most recent framework programme that is completed, which is why it was best suited for this study. FP6 was officially carried out between 2002 and 2006, but the start dates for projects in this study vary from 2003 until 2007. Because this study focuses on Germany, only projects with German project leaders are selected.

In order to receive funding for research, a research group has to consist of at least three different organizations from at least two different countries (Sixth framework programme practical guide, 2003). Projects in FP6 are divided in seven priority thematic areas. Because this study focuses on biotechnology, the first of these areas is selected: life sciences, genomics and biotechnology for health.

‘GER’

The German Federal government is actively supporting R&D efforts by public and private organizations, especially in the biotechnology sector (Broekel & Boschma, 2011), with the purpose of making it possible to reduce the time needed to transform research results into finished products (Bundesministerium fuer Bildung und Forschung – BMBF website 2012, www.era-ib.net/). Detailed records of these collaborative subsidies (or *verbundprojekten*) can be acquired on the combined subsidy website of the federal ministries (URL: www.foerderkatalog.de). There are specific regulations that the participants of these collaboration projects have to follow. For instance, the organizations have to formally agree on full cooperation on the project, letting each other make use of relevant intellectual property and know-how for free. Also the participants can use any outcomes of the project they worked on (Broekel & Graf, 2010).

To be in concordance with the goals of this research and the data from FP6, selections are made that are explained in step 3.

Networks can be built by finding the linkages between the different organizations that have collaborated in these programs. Organizations with gatekeeper positions can be designated in the two separate networks by using social network analysis. Combining these two programs as one large network provides more linkages to study, and allows for comparing three networks: the networks derived from the two separate subsidy programs, and the network entailing linkages from the combined dataset. Because overlap between the two networks likely occurs due to organizations that take part in both subsidy programs, the structure of the networks will change and new gatekeepers may be designated in the combined network.

3.1.2. assembling material for my theoretical framework

The theory chapter is structured as follows: First, some context on innovation networks is given based on literature, explaining the relation between innovation systems, network analysis and (cognitive) proximity theory. Next, cluster theory is related to these concepts, because most geographic studies on gatekeepers are rooted in cluster theory. Following this, existing literature on gatekeepers is explored to be able to form a definition and characteristics of gatekeepers for this study. The relevance of the research on innovation in emerging markets and biotechnology is treated next. Finally, some policy implications are discussed based on conclusions in existing research.

context: innovation networks

Gatekeepers are researched in a variety of different disciplines in science, and even within disciplines, the subject is defined differently. To be able to clarify how this study researches gatekeepers in a network analysis based on cognitive relations in innovation systems, a framework called *innovation networks* is constructed using literature.

innovation systems/ network analysis/ cognitive proximity

The concept of *innovation systems* corresponds with the idea of a spatial knowledge network that is beneficial for innovation. The network analysis tool can be utilized to conceptualize these innovation systems, dubbing them innovation networks. Several studies are discussed to link the concepts together. In network analysis and in innovation systems there are several different forms of proximity that might be of importance. The choice of investigating only cognitive proximity on the structural level is explained next.

cluster theory

The concept of *spatial clustering* as an innovative melting pot corresponds partially with innovation systems theory. Also, most network studies in economic geography include or are rooted in cluster theory. To be able to test the benefits of geographical proximity are not assumed in this study, and only cognitive connections between actors in the networks are explored. The concept of cluster theory in innovation network, therefore I discuss the role of gatekeepers therein in a separate box.

gatekeepers

In the paragraph about gatekeepers I chronologically consult key literature that is exemplary for relevant research that has been done on gatekeepers. Mainly, I discuss the concept of gatekeepers through several definitions and characteristics, in order of constructing a definition and characteristics of the gatekeepers for this study. Step 5 in this chapter will serve as a follow-up on network analysis in general, and gatekeepers in particular.

emerging markets and policy implications

Finally, relevant research is used to place the subject into a larger, real-world perspective, to be able to take away lessons from research results, if any. The relevance of research into emerging markets and especially into biotechnology sector is explored by discussing some (similar) studies in those fields. Next, conclusions and policy advice from similar literature is consulted to foresee possible policy implications of the possible results of this study.

3.1.3. preparing the datasets for analysis

The subsidy records from FP6 and GER are both suitable for network analysis, due to their collaboration records. To be able to use them to answer the research questions and eventually be combined into one dataset, the data from both programs need to match each other as well.

Data from both programs is selected for the biotechnology sector. In FP6 this means the data for the first thematic area is selected: life sciences, genomics and biotechnology for health. The records for the German subsidies are also selected for the biotech sector from the 'Foerderkatalog' website.

Because the FP6 officially ran from 2002 until 2006, this is the timeframe in which data is selected. The projects in the GER data are matched with the FP6 data between the first (01-11-2003) and the last (01-09-2007) start dates of the projects. These dates are matched to ensure the sharing of knowledge was indeed possible between organizations working together during that time period. Obviously, relationships may have outlasted the time period, or were already in place before FP6 started, but because this is a static, and not a dynamic study, the period in which actual collaboration was going on is taken as the research period.

FP6 is inherently an international program. From the at least three participants, at least two must be from different European Member states. The other participants can be located anywhere. For this research, only projects that have a German organization as project leader are included. The data however, contains organizations from all over the world, as participants in the German-led projects. The German subsidy data does not contain organizations located outside of Germany, but in combining the two datasets for network analysis, global linkages may be demonstrated for that program as well.

First, a framework was created in Microsoft Excel, with two different sheets for information. One for projects within the subsidy program, the second for all the different organizations and their attributes. For the projects eight variables are created:

1. Program (FP6 or GER)
2. Project number (can be matched with the data of organizations)
3. Project name (for reader reference)
4. Project funding (in Euro's)
5. Project costs (in Euro's)
6. Start date
7. End date
8. Duration (in months)

Next, sixteen columns are created for the organization sheet.

1. Program (FP6 or GER)
2. Organization name
3. Project number (for matching with the project sheet)
4. County code (geographical location)
5. City
6. Postal code
7. Country

8 - 13. Organization type (only one selected using 0 or 1)

14 - 15. Public or private (0 or 1)

16. Number of projects (the amount of projects the org. participates in)

The data of the two programs was rearranged to this format. First the data was matched. For organizations that were present in both programs, the exact names were matched. The same was done with information that was missing in one of the two. For FP6, the majority of geographical locations (city, postal code, country), and the type of organization was retrieved online, using mostly Google (URL: www.google.com), Google maps (URL: maps.google.com) and the Cordis website (URL: cordis.europa.eu) for the organization websites. The different types of organizations are: university, research organization, school, firm, public authority or miscellaneous.

The project duration and the number of projects are calculated using basic features of Microsoft Excel. To be able to make network calculations in UCINET, the excel datafiles had to be transformed into affiliation files. Because the projects represent the links between all the organizations, I created files with organizations as rows and projects as columns.

The attribute data from the excel files had to be transformed into network data, with which network analysis can be performed using UCINET 6. This software identifies the links between the organizations. Because the projects represent the links between all the organizations, I created a two-mode affiliation sheet, with organizations as rows and projects as columns. In this set up, the organizations get a ‘1’ in each project column they participate in. A one-mode datasheet, with the same actors in both rows and columns, and a‘1’when they collaborate in any project, is more common for use in network analysis because it is simpler.

In this case, inserting the data from the subsidy project records was more practical in a two-mode datasheet. Because I only need the one-mode affiliation sheets for my research purposes, and the data handling is faster and easier with one-mode data, I transformed the two-mode affiliation sheets into one-mode sheets using UCINET. The software transforms the data to the point that only organizations and their direct connections remain. The two-mode datasheet is richer in detail (dividing projects instead of just containing collaborations), but therefore also less flexible and agile in network calculations. Instead of being able to see exactly which organization takes part in which project, only the data on which organizations collaborate with each other remain after switching to a one mode sheet. For detailed discussions on this subject, see Hanneman & Riddle (2005).

3.1.4. analyze the networks FP6 (1) and GER (2) separately

First, general statistics are derived from the datasets to get to know the data. How many projects and organizations are there in each program, what is the percentage of each type of organization in each program? In how many projects does an average organization partake, in each program? In how many projects does an organization participate on average? How much of the organizations are located outside of Germany, for FP6? What percentage of organizations is located within which biotechnology cluster in Germany? How many of the organizations are located in the same counties per program? These statistics provide a general look at the data, inform about differences between the data of the two programs and organization types. This part can be done using Excel calculations.

The next step will require a switch to another type of software: UCINET 6, with which social network analysis can be performed. The data has to be prepared for, and imported into UCINET. NetDraw is a function of UCINET, with which networks can be visualized. This is a powerful tool to analyze networks and find gatekeepers with. All freely adapted descriptions for network properties are courtesy of Hanneman & Riddle (2005), unless credited otherwise.

First the two networks will be described as a whole calculating general network properties and statistics. These attributes help illustrating the networks as a whole and can tell various things about the connectivity between the nodes of the networks. Some hypotheses about gatekeepers can be deduced, which eventually help answering the research questions. The following properties will be calculated. These are averages for the whole networks. In step 5, attributes will be calculated for specific actors in the networks, in order to find gatekeepers.

Network size: The number of nodes (organizations) in the network. This is the least complicated network property, but also an important one. The following attributes have a deviating significance in networks of different sizes.

Network density: The proportion of all possible ties that are actually present. With this number, insight is given in the number of linkages in the network: the chance that two random organizations are connected in some way. The lower the density, the higher the chances for gatekeepers being present. When density is high, there's a good chance two nodes are connected in more than one way. A gatekeeper with a single connection to other nodes is less likely in a dense network. High density means that information can spread fast.

Reachability: Connected to density. Shows how much or the network is actually reachable one way or another, regardless how many other nodes fall between them. If a network has a lot of 'islands' of nodes, reachability is low. Gatekeepers are missing in that situation. If reachablity is high, however, 'pure' gatekeepers do not necessarily exist, for the same reason as when density is high. High reachablility in combination with low density however, could indicate gatekeepers because the paths exist, but there are not very many.

Point connectivity: Also connected to density. The point connectivity shows the average number of nodes that would have to be removed in order to disconnect a pair of organizations. When connectivity is low, gatekeepers are likely to exist, because they are the last ones to be taken out before an actor becomes unreachable. Again, with high connectivity, gatekeepers are likely not to exist.

Network distance: Shows the average of the shortest amount (geodesic) of steps that have to be made between two random actors in a network. A high network distance indicates a low density, and a high chance of the existence of a lot of gatekeepers. A high network distance also is an indicator that information travels slowly across the network.

3.1.5. identify the gatekeepers within each network

In step 4, general attributes were calculated for the networks as a whole, in this step a closer look will be taken at the properties of individual nodes and certain groups of nodes, in order to find actors with the role of gatekeeper in the network. The measure of *flow betweenness* will identify the gatekeepers in my network, below is explained what this measure is and how it relates to gatekeepers in this research.

In the theoretical chapter, the definition of a gatekeeper is presented*:*

An actor that holds a position in a network through which significant parts of the rest of the network are connected, and is willing to share knowledge with other actors. Removal of this actor results in significant reduction of knowledge diffusion in the network, giving the gatekeeper a potential position of power.

This actor 'keeps the gate' between large parts of its network. The more connections between other actors it makes possible, that would not be connected without it, the more important that actor is to the cohesiveness of the network. This may give an actor a certain level of power.

centrality and power

Network analysis provides various tools to calculate centrality and power measures for actors in a network. While the question of what centrality exactly is, remains subject of debate (e.g. Sabidussi, 1966; L.C. Freeman, 1979, Borgatti, 2005), but it is implied that having a central position in a network is advantageous for an actor in a network, and having a favored network position offers an actor opportunities for power (Hanneman & Riddle, 2005). Freeman (1979) dissected centrality to three commonly accepted concepts: degree, closeness and centrality. Degree is the number of nodes a certain actor is connected to, and closeness calculates how close a certain actor is to all others. With betweenness, the focus is on the connections between others that have to pass through a certain actor, for which the measure is calculated (Borgatti, 2006). *Betweenness centrality* is the measure most often used to calculate betweenness for a specific actor, and ascribes a value to the amount of times the average pair of actors need that actor to be connected to each other. This measure only calculates the shortest path between these two actors. A more comprehensive measure was introduced in L.C. Freeman et al. (1991), which is called *flow betweenness*.

flow betweenness

Flow betweenness expands on betweenness centrality, in that it assumes that actors are able to use alternative routes to other actors, which may not be the shortest, and ranks these alternative pathways based on their length (L.C. Freeman et al. 1991). Technically, flow betweenness is a centrality measure that indicates which actors are the most important mediators in the network by calculating all possible pathways between every pair of nodes that go through that mediator. I then calculates the length of the alternative paths for the same connections as if the mediator was not present. This creates a ranking of actors that are most important in connecting others, which corresponds on a technical level with my definition of a gatekeeper. The amount of organizations in the data for this study is quite large. Therefore, a chance exists that there are no 'pure' gatekeepers to be found, because every actor can be reached by more than one path. If there are 'pure' gatekeepers, that form the *only* connection between large groups of actors, they will get the highest measure of flow betweenness. When there are other paths possible between the two actors, the length of this less effective path affects the power of the 'gatekeeper' that is between the two actors.

Flow betweenness calculates the amount of flow, or information that is able to go between actors, that would not take place when the node in question would not be present, or not be willing to share. This is really the essence of any betweenness measure: the potential for withholding flow, otherwise known as gatekeeping (Borgatti et al., 2006). UCINET 6 simulates the gatekeeping process by calculating the flow between all pairs of actors with and without the actor in question. In the following equation, W is the matrix of maximum flows (paths) between nodes, $_kW$ is the matrix derived from W by deleting row and column k, and $_kW^*$ is the matrix obtained by deleting node k from the original network, and recalculating the whole flow matrix to be able to see how much that particular node is 'missed'. (Borgatti et al., 2006):

$$c_k = \sum_{i,j} \frac{{}^k w_{i,j} - {}^k w^*_{i,j}}{{}^k w_{i,j}}$$

Because the actor in that position has the opportunity to take advantage of 'it being missed' for connections between others, it is a position of power. This power becomes larger in a network position where the actor has influence in connecting more nodes, meaning that a gatekeeper for one pair of actors adds less to the cohesiveness of the whole network than an actor that is important for the connection of lots of pairs. In short, flow betweenness will calculate for every actor the extent to which it holds a position in a network through which significant parts of the rest of the network are connected. This way, a ranking of flow betweenness results in a ranking of potential 'gatekeeperness' of actors. This, in combination with the willingness to share knowledge and the characteristics set out in the theoretical chapter, leads to finding the gatekeepers in the network.

nFlow betweenness

To find the actors in my networks that are gatekeepers according to the definition set in paragraphs 2.2.2 and 3.1.5, first the flow betweenness is calculated in UCINET 6. All actors will be listed from highest to lowest flow betweenness. With UCINET, I also calculate the percentage of betweenness in ratio to the total amount of actors, called *nFlow betweenness*. This measure is referred to, to be able to compare the scores of both subsidy program networks, and potentially be able to compare to other networks in similar studies.

Because there is no point of reference for me to which actors with which characteristics will prove to be actual gatekeepers, an arbitrary guideline is based on the results of this network. After calculating the nFlow betweenness, both networks showed a limited amount of organizations with an nFlow betweennes above 1%. Picking the first five or ten organizations from the list of flow betweenness is also an option, but this way, a clear line is drawn in the amount of *gatekeeperness* based on the value that I measure.

analysis of the gatekeepers

Unfortunately there is no way of knowing how these levels relate to actual *gatekeeperness*, especially in relation to other existing networks. In less dense networks, for example, there might be a lot more actors with nFlow betweenness above 1%. That is why this '1% measure' is applied with caution, until it is studied more thoroughly, in a variation of networks. That is beyond the goals of this study, but can lead to a fruitful continuation for other research. In this research, the actors with a nFlow betweenness above 1% will be referred to as gatekeepers because they fit the technical definition set by this study.

For both the subsidy programs, the gatekeeper actors are analyzed more closely, in terms of organization type and geographic location. They are also compared to each other and to the average statistics of the organizations their programs as a whole. To determine if the actors with the highest flow betweenness are actual gatekeepers, the characteristics and a qualitative analysis are performed on these actors in, explained in paragraph 3.2. and executed in paragraph 4.5.

gatekeepers in other studies

Other studies that are concerned with identifying gatekeepers usually do so from a perspective of primary data, like interviews and company records. These researchers find gatekeepers on an individual level, usually in a case study on a cluster or industrial district. Giuliani & Bell (2005) find, through interviews, a number of different indicators for gatekeepers, like absorptive capacity and of course the intra-cluster connections of organizations. Morrison (2008) identified district leaders and gatekeepers through interviews with members of an industrial district, determining their role in the network by analyzing relations with other actors, codification efforts and production details. Leaders and gatekeepers are depicted in a network after they were identified through these interviews, using the network analysis more as a descriptive tool than an investigative one.

Kauffeld Monz & Fritsch use network analysis to identify gatekeepers, much like in this study. In a network comprised of information from a large scale survey in 18 regions, network actor characteristics degree, betweenness and brokerage are calculated and used as indicators for gatekeepers, in combination with indicators for knowledge absorption and transfer capacity and innovation measures. While this approach is similar to the methods of the present study, using only flow betweenness offers a fitting measure for the exact definition of a gatekeeper for the goals of this research. Using a single calculation which denotes the cohesiveness of the whole network with and without every actor, makes the calculation and interpretation process simple and effective.

Finally, the study of Malerba et al. (2006) also utilizes network analysis to find hubs and gatekeepers. They use a semantically different definition of a gatekeeper, designating actors that broker within and between networks as hubs and gatekeepers for practical reasons. Hubs and gatekeepers are, like in the present study, defined by a certain network position, whose characteristics may have to be determined, instead of met.

3.1.6. combine the two networks through step 4 and 5

In this step, I combine the FP6 program and the GER program data in the same network file in UCINET. With the complete dataset, new linkages may arise, combining the two programs into one large network. Connections between the two dataset exist because of organizations that are active in both subsidy programs. With this combined dataset, three different networks are available for analysis: the two separate programs and the combined network.

The combination is done in Excel, before the network is imported in UCINET. In combining the networks, the nodes that are active in both programs will show up as doubles. These duplicates are removed, to be able to describe the new network the way the separate program networks were also described.

The combined network may show similar or different results from the separate networks. By combining the datasets, the information retrieved from analysis may be validated for a larger population. By analyzing the similarities and differences between the three networks, the reliability of these different networks may be determined. I assume that the results from the combined network are averages between the two networks. In this case, it can be assumed that the two separate networks are already reliable sources of information for research into innovation networks in the biotechnology industry. If the network characteristics change significantly after the combination of the two programs, the new and larger research population may be a more reliable source than the separate programs. This depends on a number of variables, in which the organizations that participate in both programs may play a large role. They are then likely to be the gatekeepers, making the connections between the two networks possible.

By repeating steps 4 and 5 for the combined network, a thorough analysis will be possible for the combined network. A new set of characteristics and gatekeepers may arise. In the next step, these differences are analyzed.

3.1.7. compare the gatekeepers from the combined network (3) with the identified gatekeepers from network 1 and network 2

In this step, I compare the gatekeepers based on nFlow betweenness from the three different networks, and the organizations that have an nFlow betweenness above 1% are compared,

The gatekeepers that emerge from the combined network are then analyzed. First, it is interesting to see if there are actors that we can call gatekeepers in the combined network, that were also gatekeepers in the separate networks. Then the division of type of organization is described in the combined network, which is compared to the numbers of the two separate program networks. An analysis of the number of projects in which the gatekeepers participate on average is also compared to the same numbers from both separate programs.

After this some geographical information about the gatekeepers in the combined network is analyzed. How many of the gatekeepers are German? I match the geographical locations of the gatekeeper organization with the biotechnological clusters in Germany, in comparison to the separate subsidy programs.

A regression and correlation calculation is performed between the network of cognitive links, and the network of geographical links. To do this, the affiliation sheet of the cognitive links (the collaboration projects) is matched with a new one, in which the geographical locations of the organizations form the connections. The new affiliation sheet is created by matching all organizations that are located in the same county. The regression and correlation between these affiliation sheets are calculated using UCINET, indicating the chance that two organizations that are collaborating in a project are also located in the same area. If this chance is high, the possibility of geography playing a role in these networks is higher as well.

The goal of the next step is to find out if the gatekeepers in this network naturally act as a gatekeeper for geographical entities as well, without any assumption that a geographical cluster is present.

Now that the gatekeepers are identified, they can be analyzed on the node level. The ego network of an actor is the network that surrounds it, be it by one step (the direct connections), or more steps. The point of view of an ego network is focused on one actor and its surrounding network. The geographical details of the ego networks of the gatekeepers are analyzed, in order of finding out if there are geographical patterns. To how many actors is the gatekeeper connected that are located in the same region, and do these actors rely on the gatekeeper for their inter-regional connections? Or do they have outside connections as well? The data provides county codes for the actors. This is a somewhat crude measure because counties have different sizes. After the gatekeepers have been identified, the counties of Tübingen and Reutlingen have been merged to be more similar in size to the counties in which the other gatekeepers reside. Also, the gatekeepers identified in those counties have been found to have a corresponding and intensive regional focus. This part of the study is a first step into the role of geographical proximity in innovation networks. It should be noted that there are other possibilities for organizations to be connected, other than through collaboration in a German or European subsidy program, so this study will take a peek at the role of geography in this network as a first step. A thorough analysis of the role of geography in innovation networks is beyond this research, because intensive primary data has to be collected in the ego networks of the identified gatekeepers.

Netdraw (Borgatti, 2002) is a function of UCINET with which networks can be visualized with dots and lines. For smaller networks it is a very useful tool, because a visualized network can be of aid in describing certain network features. My network however, is too big to visualize. Actors will be indistinctly close to each other and the many lines will obstruct their view. Now that I have arrived at a small group of actors of interest, some Netdraw visualizations may be of interest, for example the ego networks of found gatekeepers, and the links between the group of gatekeepers. An ego network is the network around an actor of interest, for example every actor that is directly connected to that node. For some gatekeepers this might result in an interesting visual, to see how it acts as gatekeeper for different actors around it.

3.2. qualitative approach of research question 3

With the gatekeepers identified technically, in this step I try to find out whether these actors actually correspond to the characteristics of gatekeepers set out in the theoretical chapter. This step is also designed to be able to answer the third research question about the general application of the results of this study, and its possible policy implications.

4. *Are the results generally applicable in innovation networks of emerging technology markets and what are possible policy implications?*

Following the results of research questions 1 and 2, an overall comparison is made between the gatekeepers that are found and analyzed, and their relation to the theory on innovation networks and the role of geography herein.

To be able to form conclusions that are useful for organizations and policymakers, a framework will be built that is based on the first two research questions. The third question helps to place everything that was learned about gatekeepers in a real world setting.

Of the gatekeepers that have been identified in the previous paragraphs, key organizations will be chosen that represent every major type of gatekeeper of the group. There is no need to qualitatively analyze two gatekeepers that are similar in every way, but the number of analyzed actors in this paragraph depends on the variety of types of gatekeepers that emerge. With types one can think of the type of organization, whether it is public or private, where it is located and what size it has in terms of links in the network.

The key types of identified gatekeepers are treated to a confined study, to give the technically indentified gatekeepers some real world background. This will help placing them in a framework on which further studies can continue. The theoretical chapter provides guidelines for conclusions and previous policy implications, while the cumulative results form the basis for new conclusions and possible new policy implications.

The chosen gatekeepers are described by the following points retrieved from public sources like their official websites, their annual reports and catalog websites:

- The type of organization, according to the datasets for this study and the official websites
- The goals and mission of the organization
- The number of employees
- The budget according to the annual report, and if available, the division of public and private funding
- Any information relating to official partnerships with other organizations
- The location of the organization and its relevance, also in relation to other organizations that are geographically close, or in one of the biotechnology clusters.

3.3. chapter summary

In this chapter the method and order of actions for this study were explained. A step by step plan was laid out, which is the guidance for the next chapter. Also, this chapter serves as an explanation and reference for choices made.

4. results

introduction

In this chapter, I will execute steps 4, 5, 6 and 7 as described and explained in the methodology chapter. By taking these steps, the gatekeepers in the networks at hand can be found and analyzed. The steps are:

- step 4: Analyze the networks FP6 (1) and GER (2) separately
- step 5: Identify the gatekeepers within each network
- step 6: Combine the two networks through step 4 and 5
- step 7: Compare the gatekeepers from the combined network (3) with the identified gatekeepers from network 1 and network 2

I will treat each of these steps in separate paragraphs, followed by a paragraph about the qualitative approach for research question 3.

4.1. step 4: analyze the networks FP6 and GER separately

In this step, I will analyze the two programs using (1) statistics and (2) network analysis. This will be done extensively in order to get a better understanding of the conditions in which gatekeepers operate. Firstly, the data will be taken apart, in order to explore and describe the different attributes of the two programs (i.e. amount of organizations per program, division of organization types, public and private, division of organizations in projects and projects in programs). This will help understand the construction and patterns of the two programs, and will eventually help explain the conditions in which gatekeepers exist. Secondly, utilizing network analysis, relationships and patterns between the organizations will be shown from a wide array of angles, to be able to understand where the gatekeepers are located in the networks, and what their network characteristics are. This will provide the groundwork for finding and explaining the gatekeepers in the networks, which will be done in step 5.

general statistics

Table 4.1 shows the size of the programs after selecting the data for this study (step 3). In the network, the organizations are the nodes and the projects form the links between them. Also shown for both programs, are the average size of projects, as well as the number of projects the average organization takes part in. This paragraph is ordered in the same consecution: first the data is discussed from the project perspective (the links), next the organization will be the point of view (the nodes).

	# unique organizations	# projects	$\bar{x}$ orgs/project	$\bar{x}$ projects/org
FP6	726	103	14	2
GER	286	200	3,4	2,4

table 4.1: number of organizations and projects in the two programs

Between programs, in the same time period, there are considerable differences. The Sixth Framework Programme has a lot more organizations and less projects than the German program. The amount of participating organizations that meet the restrictions for this research can be explained by the scale of the programs. FP6 is an international program, which may contain organizations from all over the world, while the German program only allows German organizations. Because the projects form the links between the different organizations, I could expect the organizations in the Framework Programme to be more densely connected than the organizations in the German program, because there are more organizations that participate in less projects. The exact statistics on this can be found in table 4.2.

	FP6	GER
Max	146	17
Min	2	2
Mean	14,0	3,4
Standard deviation	15,7	2,3
Median	9	3
Skewness	6,1	3,0

table 4.2: number of organizations per project in the two programs

The average project in FP6 contains 14 organizations, considerably more than the 3,4 organizations of the German program. The statistics above show that the organizations in FP6 are distributed unevenly across projects, with a high standard deviation, which shows that on average, the number of projects vary almost 16 organizations from the average amount. Also calculated is a high skewness score and a much lower median in relation to the mean. The amount of organizations per project in the German program are also skewed, but to a lesser extent.

Table 4.3 displays average percentages of types of organizations per project and their standard deviations. These statistics are found by separately generating percentages for every single project, and calculating the means and standard deviations for these means.

	FP6 (%/project)	FP6 Std	GER (%/project)	GER Std
University	44,1	21,4	38,1	31,2
Research Organization	30,6	19,6	21	28,0
School	0,3	1,3	0,1	1,7
Firm	19,3	17,2	39,4	34,2
Public Authority	4,1	7,0	0,2	2,3
Miscellaneous	1,6	6,6	1,2	7,3
Total	100		100	
Public organization	72,1	18,6	44,5	32,8
Private Organization	27,9	18,6	55,5	32,8
Total	100		100	

table 4.3: type of organization distribution of average project

The average project shows a clear division between universities and firms in both programs. Projects in FP6 on average contain a majority of universities, less research organizations and even less firms, each about 10% lower than the one before. Within the German program, firms and universities both make up around 39% of each project, and research organizations are at a little more than half that percentage. All the means have quite substantial standard deviations, so the actual distribution in the projects can vary significantly. The importance of these organization types cannot be derived from these numbers, but they do show that a typical project is likely to embody at least a university, a research organization and a firm. Schools, public authorities and miscellaneous organizations are scarcely present. This is not surprising, since they are not likely to directly participate in biotech innovation (for reference see chapters 2 and 3).

Table 4.4 shows the percentages of types of organizations that are present in both programs. Table 4.3 showed the percentage buildup of projects, in which organizations that take part in more than one project are counted multiple times. Table 4.4 takes in account percentages of unique organizations in the complete dataset. There are some notable differences.

	FP6 (%)	GER (%)
University	37,6	21,7
Research Organization	29,2	12,2
School	0,3	0,3
Firm	26,6	63,3
Public Authority	4,5	0,3
Miscellaneous	1,8	2,1
Total	100	100
Public organization	65,2	31,5
Private Organization	34,8	68,5
Total	100	100

table 4.4: percentages division of organizations in the two programs

Most organizations in FP6 are universities, like in table 4.3. Research organizations and firms are second and third. The differences in percentages show that universities and research organizations are more often participant in more than one project than firms. This suggests that research organizations are more often the ones that connect different projects. This is especially visible in the German program, where a large majority of unique organizations are revealed to be firms, while table 4.3 shows that the ratio of organizations in projects is much more evenly spread. Firms participate in a lower number of projects than universities and research organizations. This could be caused by size differences of these types of organizations, which is discussed in the theory (chapter 2) and discussion (chapter 6).

Like determined in table 4.3, schools, public authorities and other organizations do not play a considerable role in either program. The ratio of the number of public and private organizations is reversed in the programs. This was already visible in the organizations types above, but here the division demonstrates that the programs are clear opposites.

Taking a look at the typical organization in each program, the differences between the two programs are decidedly smaller, as can be seen in table 4.5.

	FP6	*GER*
Max	*24*	*37*
Min	*1*	*1*
Mean	*2,0*	*2,4*
Standard deviation	*2,3*	*3,4*
Median	*1*	*1*
Skewness	*4,0*	*5,4*

table 4.5: number of projects that organizations participate in

The distribution of organizations in programs is also skewed, organizations in FP6 take part in less projects than organizations in the German program, which was to be expected from the tables 4.3 and 4.4. 69% of the organizations participate in only one project in FP6, in the German program this is 63,3%. These organization cannot be gatekeepers because they do not form the links between the different projects.

network analysis

The two programs will be described using general network characteristics calculated with UCINET (Borgatti et al., 2002). The results are compared to the comments from the methodology chapter. In that order, network size, network density (including reachability and connectivity) and network distance will be described.

Network size: The size of the network is the sum of all the nodes, which are in this case the unique organizations. FP6 has 726 organizations, GER has 286. This is already discussed above.

Network density: The network density shows the percentage of ties that are present in the network. 100% would mean that every node is directly connected to every other node. Table 4.6 displays the density of both networks. Reachability and connectivity are also forms of density, which will be discussed below.

FP6	*.084 (8,4%)*
GER	*.033 (3,3%)*

table 4.6: network density

These percentages both seem quite low. The first reason of this is that these are large networks. The larger a network, the large the chance that the density is low, because direct connection with every other node are less likely to exist. Therefore, reachability will provide a better measure for interpreting the connectedness of these networks. The density figures also depend on the type of data, which is the second reason. The subsidy group structure of the network makes a high density improbable: the only direct linkages in our programs result from project groups. Most organizations only participate in one project, the other participants are then the only direct linkages they have. This fragmented structure can easily result in low densities.

The FP6 has a higher density than GER, as predicted in the general statistics. As explained in the methodology chapter, low density could facilitate the existence of gatekeepers, because there is a high chance that pairs of nodes are connected in only one way.

Reachablity: The network reachability traces every path between every pair of nodes. The result is a percentage of possible paths between any pair of nodes. Table 4.7 shows these percentages for both networks.

FP6	*.989 (98,9%)*
GER	*.619 (61,9%)*

table 4.7: network reachability

There is a large difference in reachability in the two networks. In FP6, almost all points are connected to each other in some way, while in the German program 38,1% of ties cannot be made. This seems to affirm the organization/project proportion observed in the general statistics. In the German program, groups of nodes (projects) are sometimes not connected, which means that a gatekeeper is missing. The high reachability of FP6 however, could indicate the existence of gatekeepers in combination with the low density level, as explained in chapter 3. The same goes for GER because a majority of nodes can still reach each other.

Connectivity: The point connectivity shows the average number of nodes that would have to be removed in order to disconnect a pair of organizations. The results are shown in table 4.8.

	FP6	*GER*
Max	*245*	*41*
Min	*0*	*0*
Mean	*16,97*	*1,63*
Standard deviation	*30,25*	*2,98*

table 4.8: network point connectivity

Again, there is a large difference between the two programs. In this case it has to do with network size as well as the other two density attributes. Because FP6 is a larger network, more paths between actors would be possible than in a smaller network with a similar density. In this case, GER is smaller and has a lower density and reachability. Point connectivity, could give an indication on the possibility of gatekeepers. FP6 connectivity has a high mean, but also a high standard deviation. This does not indicate a lot of gatekeepers, but it also does not indicate that they do not exist. The low point connectivity for GER could indicate a high number of gatekeepers, because the network is generally weakly connected, but there is also a high amount of pairs that is not connected at all.

A frequency query in UCINET shows the frequencies of the number of nodes that would have to be taken out for a disconnect. If there is one node to be taken away for two others to be disconnected, that node would be a gatekeeper for that pair of nodes. There are 48950 of 'one' nodes present in FP6 that are the only link between two others. There are 24892 in GER. To explain why these numbers are so high, for example consider that in a string of single nodes, every one of those would be a gatekeeper to the two nodes surrounding it. They would not be indicated as gatekeepers due to a lack of centrality, which will be treated in paragraph 4.2. The frequency query also shows how many nodes are already absent in connecting other actors. These non-existent nodes take the first spot in the GER network with 31022, while FP6 only has 5776, the 23 spot in the list of frequencies. The reachability already showed that there is a large difference in these numbers.

network distance: The average geodesic distances of both networks are displayed in table 4.9. The geodesic distance is the shortest possible path between a pair of nodes.

	FP6	*GER*
Max	*7*	*8*
Min	*1*	*1*
Mean	*2,4*	*4,9*
Standard deviation	*0,9*	*2,5*

table 4.9: network geodesic distance

The distances between nodes in the FP6 network are quite low. Every pair of actors that is able to reach each other is able to do so in a maximum of 7 steps. This is not a low maximum, but the average pair of nodes can reach each other in 2,4 steps. The German program network has more distance. The maximum is only one step higher, but the average is almost 5 steps between nodes. However, the standard deviation is also higher, so there is more variation in the network. A high distance could facilitate gatekeepers, so the GER network has a higher chance of having gatekeepers.From the statistics and network characteristics calculated in this paragraph, it could be carefully expected that there are more gatekeepers to be found in the German program network. In GER there are less actors in more projects, while those actors are less active in more than one project than in FP6. Also, the GER network is less densely connected, there is less reachability and connectivity and more distance. These are all conditions an actor in the right position could benefit from. This will be addressed in the next paragraph, in which the gatekeepers will be pointed out.

4.2. step 5: identify the gatekeepers within each network

The centrality and power of all organizations will be quantified in this step, by determining the flow betweenness for both networks. Flow betweenness is a centrality measure that indicates which actors are the most important mediators in the network by calculating all possible pathways between every pair of nodes, as well as the length of the alternative paths for the same connection. For each actor, the measure reflects the number of times the actor is in a flow (any flow) between all other pairs of actors. The most indispensible nodes get the most flow betweenness. To compare the two networks, nFlowbetweenness is calculated, which is the flow betweenness relative to the size of the network. Table 4.10 displays the results of calculating nflow betweenness for both networks.

	FP6	*GER*
Max	*3,615*	*6,951*
Min	*.001*	*0*
Mean	*.157*	*.297*
Standard deviation	*.269*	*.786*

table 4.10: nflow betweenness descriptives

In both networks some actors are evidently more central than others, there is a rather large relative variability in both networks. The standard deviation of each network is substantially higher than the mean, so there is a lot of variance in flow betweenness among the actors of the networks.

Tables 4.11 and 4.12 show the gatekeepers in nFlow betweenness for both networks. The actors that have an nFlow betweenness that is higher than 1 percent are designated gatekeepers, as explained in chapter 3. Network analysis in the previous paragraph indicated the existence of gatekeepers in both networks. The following represent the top selection of actor-gatekeeperness.

	Organization	*FlowBet*	*nFlowBet*
1.	Charité - Universitätsmedizin Berlin	18975,2	3,615
2.	Deutsches Rheuma-Forschungszentrum Berlin (DRFZ)	13458,0	2,564
3.	Steinbeis Forschungs- und Entwicklungszenteren GmbH	13268,1	2,528
4.	Max Planck Gesellschaft für die Wissenschaften e.V. (MPG)	12862,2	2,450
5.	Archimedes Foundation	12424,6	2,367
6.	Institut National de la Santé et de la Recherche Medicale	8117,7	1,547
7.	The University of Oxford	5641,3	1,075

Table 4.11: flow- and nflow betweenness centrality for FP6

	Organization	*FlowBet*	*nFlowBet*
1.	Fraunhofer Gesellschaft e.V. (FhG)	5626,5	6,951
2.	NMI an der Universität Tübingen	5265,2	6,505
3.	Westfälische Wilhems-Universität Münster	3407,3	4,210
4.	Max Planck Gesellschaft für die Wissenschaften e.V. (MPG)	3404,9	4,207
5.	Eberhard Karls Universität Tübingen	3066,2	3,788
6.	Humboldt Universität zu Berlin	2638,0	3,259
7.	Medizinische Hochschule Hannover	2235,0	2,761
8.	Universität Stuttgart	2075,4	2,564
9.	Charité - Universitätsmedizin Berlin	1949,5	2,409
10.	Universität Rostock	1479,9	1,828
11.	Technische Universität München	1353,7	1,672
12.	Freie Universität Berlin	1145,9	1,416
13.	Technische Universität Carolo Wilhelmina zu Braunschweig	995,4	1,230
14.	Johannes Gutenberg Universität Mainz	938,8	1,160
15.	Ruhr Universität Bochum	917,0	1,133
16.	Christian Albrechts Universität zu Kiel	878,3	1,085

Table 4.12: flow- and nflow betweenness centrality for GER

Like expected, the network of the German program has more organizations in a gatekeeper position. All the organizations in both subsidy programs are universities or research organizations. In FP6 there are 7 organizations with this qualification, in GER there are 16. In FP6, 5 of the 7 (71,4%) gatekeepers are research organizations. In the German program only 2 out of 16 (12,5%) are research organizations. Only one organization is not public: The

Steinbeis foundation is a non-for-profit but privately managed foundation. There are no firms, the first company on the organizations list of flow betweenness are number 17 in FP6 and number 23 in GER. Two organizations are gatekeepers in both networks, the Max Planck research institute and the Charité medical university in Berlin.

Research organizations and universities have the creating, sharing and usually spreading of knowledge as a goal. The fact that every gatekeeper organization has a public mission (even Steinbeis, see paragraph 4.5), affirms the notion that the willingness to share information is probably an important characteristic of a gatekeeper.

Of these gatekeepers, 4 organizations are together in Berlin, 2 in Stuttgart, 2 in Hannover and 2 in München, all locations that are designated clusters for biotechnological activity (see theoretical chapter). Three gatekeepers in FP6 are located outside of Germany: The French medical institute for public health, Oxford university and the Archimedes Foundation in Estonia.

Interestingly, the Archimedes foundation in FP6 is an non-German organization that only takes part in two projects of moderate size (both 9 organizations per project, the median for FP6). It is notable that these projects both feature mostly non-German organizations. The Archimedes foundation apparently is an important link in connecting the FP6 network as a whole. Similarly, in GER the Humboldt university only participates in three small projects of 4, 4 and 2 organizations. This is an indicator that an actor does not have to have a lot of direct connections in a lot of different projects to be a gatekeeper.

Statistics on the number of projects in which these gatekeepers participate are listed in table 4.13

	FP6	*FPGATE*	*GER*	*GERGATE*
Max	*146*	*20*	*17*	*45*
Min	*2*	*2*	*2*	*3*
Mean	*14,0*	*11,0*	*3,4*	*15,6*
Standard deviation	*15,7*	*8,1*	*2,3*	*10,7*
Median	*9*	*11*	*3*	*13*

table 4.13: organizations per project

In FP6, the gatekeepers participate on average in less projects than is the mean for all organizations in the program, while this number is higher by a large degree in the German program. Note that in the Framework Programme the medians are closer together, and that the program has a very skewed distribution of number of participants per project. The medians in GER do vary a great deal however, in this program most gatekeepers take part in an unusually high number of projects, connecting project groups to each other. The

Humboldt university has the lowest number of projects, which is still very near the program average. This again indicates that being part of a large number of projects does not necessarily make an organizations a gatekeeper.

In paragraph 4.1 it was expected that the German network, in the way it is structured, probably features more gatekeepers. This step confirmed that expectation. It is also discovered, that the size of the projects in which gatekeepers are active is not decisive, as well as the number of projects an organization participates in. In the next step, the two programs will be combined to examine overlap and differences, and to learn more about the gatekeepers that are found in the new dataset.

4.3. step 6: combine the two networks through step 4 and 5

In this paragraph, the two programs are combined into one new dataset, to be able to investigate to what extent the two programs overlap, and what is new. Some of the statistics and most of the network characteristics will also be repeated an compared to the two datasets. Some of these number are left out because they do not provide new information. It is also possible to investigate some geograpical patterns with the combined dataset. This will be done at the end of this paragraph. In the next paragraph, new gatekeepers will be determined based on this new dataset.

In combining the two datasets, 938 unique organizations were left, which means that 74 organizations are active in both programs. These are 74 links between the two datasets, through which UCINET will calculate new paths between nodes in the much larger network.

general statistics

Most of the simple statistics will not change in the combined data, they will be an exact average between the old numbers. The organizations per project for exampe will not change, because projects are program related. Because some organizations were present in both programs, the division of unique organizations per type has changed in the combined data. However, these statistics do not provide for an interesting analysis because the differences are very small.

network analysis

Table 4.14 displays all of the network characteristics that were calculated for both programs, in comparison with the same numbers for the combined dataset.

	FP6	GER	Combined
Size	726	286	938
Density	8,4%	3,3%	5,3%
Reachability	98,9%	61,9%	88,6%
Connectivity	16,97	1,63	11.39
Standard deviation	30,25	2,98	24,93
Distance mean	2,4	4,9	3,1
Standard deviation	0,9	2,5	1,9

table 4.14: Network characteristics compared

The network size is 74 nodes smaller than the combined dataset, which is explained above.

The density of the new network almost is an average of the two programs. This was to be expected, because density only deals with direct connections. In combining the data, new possiblities for ties between nodes originate, while only the 74 crossovers actually exist. In a network in which density is already very low, this results in a combined density that is just below the average of the two programs.

The reachablity is higher than the average of the other two programs. This means that in the combination, more paths have been created that did not exist before. The German program has gained paths from the combination through the Sixth Framework Programme.

The connectivity also went up a great deal for GER, there are now on 11,39 nodes that have to be removed in order to disconnect an average pair of nodes. The standard deviation is, like in FP6, very high, so there is a lot of variability in these numbers.

The average distance between two nodes is lower than the average of the programs, which is 3,7 steps. This means, like in the characteristics above, the combination has made the network connected more effectively. This will also probably mean that there will be less gatekeepers in the new combined network. Simply put, with more effectively connected nodes, organizations will have less trouble connecting to others, which gives less organizations a position of power. In the next paragraph, these gatekeepers will be tracked down and analyzed.

4.4. step 7: compare the gatekeepers from the combined network (3) with the identified gatekeepers from network 1 and network 2

The flow betweenness for the combined dataset has been calculated, the results are in table 4.15. nFlow betweenness is again the measure with which the different dataset can be compared.

	FP6	GER	Combined
Max	3,615	6,951	3,500
Min	.001	0	0
Mean	.157	.297	.112
Standard deviation	.269	.786	.264

table 4.15: nFlow betweenness descriptive combined

The number of organizations in the combined dataset naturally is a lot longer. The mean of the relative flow betweenness per organization is lower than in both separate programs. This means that there are on average less nodes in the network that are important to the cohesiveness of the network, so the network is stronger and less dependent on gatekeepers. This is also apparent when taken into account the maximum measure found in the combined data, which is lower than both of the first two gatekeepers in the two separate networks.

Table 4.16 displays the list of gatekeepers with an nFlow betweenness higher than 1%.

	Organization	FlowBet	nFlowBet
1.	Charité - Universitätsmedizin Berlin	30761,8	3,500
2.	Max Planck Gesellschaft für die Wissenschaften e.V. (MPG)	30604,3	3,482
3.	Fraunhofer Gesellschaft e.V. (FhG)	29223,8	3,325
4.	NMI an der Universität Tübingen	21658,0	2,464
5.	Deutsches Rheuma-Forschungszentrum Berlin (DRFZ)	16846,0	1,917
6.	Eberhard Karls Universität Tübingen	16362,7	1,862
7.	Archimedes Foundation	15059,7	1,713
8.	Steinbeis Forschungs- und Entwicklungszenteren GmbH	14203,7	1,616
9.	Medizinische Hochschule Hannover	9979,8	1,135

table 4.16: gatekeepers in the combined network

As expected, in comparison to the two separate networks, the number of gatekeepers in this network is limited. FP6 had 7 gatekeepers, GER had 16. This was expected because it is an effectively organized network according to the characteristics calculated in the previous paragraph. Again, only universities and research organizations are gatekeepers, once more affirming the probable importance of the willingness to share knowledge. 44,44% are universities, 55,56% are research organizations. One organizations is not public: the non-for-profit Steinbeis research center. Numbers 1 and 2 are exactly the organizations that already were gatekeepers in the separate networks, they are also the organizations with the highest number of projects, as can be seen in table 4.17. there is still one non-German gatekeeper from FP6, the Estonian Archimedes Foundation connects parts of the network that would be hard to reach otherwise. There is a chance that a number of direct connections of the Archimedes foundation are not present because this dataset is focused on German led projects, and that this is the reason it technically emerges as a gatekeeper for the organizations it connects to the rest of the network.

Table 4.17 displays the organizations per project among the gatekeepers in comparison to FP6 and GER.

	FPGATE	*GERGATE*	*Combined*
Max	*20*	*45*	*65*
Min	*2*	*3*	*2*
Mean	*11,0*	*15,6*	*20,7*
Standard deviation	*8,1*	*10,7*	*19,9*
Median	*11*	*13*	*14*

table 4.17: organizations per project

The average amount of projects of these gatekeepers is higher than in both of the programs. This could be expected, because the list is a combination of the two programs, with the overlap explaining the increase. The Max Planck society (65) and the Charité medical university (31) take part in the highest amount of projects of the whole dataset. However there is still room for the two small projects of the Archimedes foundation and the four project in which the Steinbeis organizations participates (3, 5, 33 and 9 participants, so one very large one). The number of projects by itself is not a decisive measure.

geographical patterns

In order to discover if the project based linkages between the organizations in our combined network also have a geographical dimension, the network has been simplified to just the German organizations. There are 374 German organizations in this network, which can be used for location analysis.

According to Broekel (2011), there are 7 main biotechnology clusters in Germany. These clusters have been compared to the amount of organizations in those areas in the combined network and can be found in table 4.18

Berlin	*46*
Göttingen	*10*
Hamburg	*19*
Munich	*25*
Rhine-Main	*17*
The Rhineland	*26*
Rhine-Neckar	*26*

Table 4.18 frequencies of organizations in biotechnology clusters

169 of 374 (45,2%) organizations are in these clusters. There are 127 counties in the dataset. On average, each county has 3,6 organizations. Table 4.18 shows that in each of these biotechnology clusters the number of organizations in our data is much higher.

To be able to statistically compare the locations of the organizations with the links between the organizations, some regression and correlation tests have been carried out using UCINET. This determines if organizations that are located in the same county participate in the same projects more often than organizations that are not geographically close to each other. Two networks have been compared: the affiliation network with projects as links, and a network with organizations and county codes and links, the results can be found in table 4.19.

	Regression	*Correlation*
Coefficient	*.001*	*.0301*
Significance	*.002*	*.0016*
Unstandardized coefficient	*.055*	-

table 4.19: statistical differences between cognitive and geographic networks

Network regression shows the nature and direction of a relation between to variables. The R^2 is .001, with a probability of .002., which means that there is a very small but significant positive relation between organizations taking part in the same project and being located in the same county. The unstandardized coefficient of .055 then means that if two organizations are in the same county, this increases the probability of them participating in the same project by .055%.

With network correlation the strength between the similarities of two networks can be measured: The test results in .0301 Pearson's correlation, with a significance of .0016. There is a very weak but significant relationship between the cognitive (project-linked) network and the geographically linked network. With large datasets like these, significance can be unreliable with Pearson's correlation, but the small correlation figure in itself tells that the effect of geography on collaboration is very low.

The geographic tests for regression and correlation show that there is a significant positive relation between geographical location and project participation that is very small. This basically says that the chance two participating organizations are likely not located in the same county. A more comprehensive study should be carried out to find answers to the role of geography and collaboration between these organizations.

All of the gatekeepers in the combined network are located in one of the seven biotechnology clusters in Germany. Three organizations are located in Rhine-Neckar, two in Berlin, two in Hannover and one in München, which only leaves the Archimedes foundation from Talinn.

An analysis of the ego networks of each of the gatekeepers yields some interesting results. The direct connections of the gatekeepers have been analyzed based on the region in which they are located, to find out if the gatekeepers also have a geographical role, which is often suggested by economic geographers. If a gatekeeper in the network also ‘keeps the gate’ of their geographical region, it should be the only actor located in its ego network that has access to sources from outside that region. It should be noted that there are other possibilities for organizations to be connected, other than through collaboration in a German or European subsidy program, so this study will take a peek at the role of geography in this network as a first step.

Also, these results are based on the county codes, which is a somewhat crude measure. The counties of the nine identified gatekeepers have been studied, and the counties of Tübingen and Reutlingen have been found to be geographically and cognitively connected. Therefore, these have been merged to be able to compare them to the other counties in which gatekeepers are found. The organizations that have been found to be connected exclusively to organizations in their own region, and for which the identified gatekeeper acts like a geographical gatekeeper as well, all only participate in the German subsidy program. This is because every FP6 project had at least one organization from another country in it, and generally contains larger project groups than GER.

Charité - Universitätsmedizin Berlin

The Berlin region, with 36 organizations, is the largest region in the network. Two of the gatekeepers are located there, Charité and the DRFZ. In the ego network of Charité are 276 organizations, the largest of the network. 4,35% of these directly connected actors are located in Berlin. These 12 organizations all have multiple connections outside of Berlin, so the Charité is not a geographical gatekeeper for these organizations.

Max Planck Gesellschaft für die Wissenschaften e.V. (MPG)

The München region has 20 organizations in the combined network. Two gatekeepers are located in München, Max Planck and the Fraunhofer community. In the ego network of Max Planck are 279 organizations, 9,32% of these are located in the München region. However, most of these are important players themselves, like Fraunhofer and the technical university of München. All but one of them (the Blutspendedienst des Bayerischen Roten Kreuzes GmbH) have connections outside of the region as well. Except for this single organization, none of the organizations in the relies on the Max Planck community for inter-regional connections.

Fraunhofer Gesellschaft e.V. (FhG)

In the ego network of Fraunhofer are 123 organizations, 5,69% of these are located in München, but all have connections outside of the region.

NMI an der Universität Tübingen

Twenty organizations from the region of Tübingen are present in the combined network. As explained in the methodology, Tübingen has been merged with Reutlingen, because it can be assumed that this is one regional network. Two gatekeepers are located there, the NMI and the Eberhard Karls university. In the ego network of NMI are 17 organizations, 29,41% of these are located in the Tübingen region, all have connections outside of the region. The NMI itself is only active in projects of the GER program.

Deutsches Rheuma-Forschungszentrum Berlin (DRFZ)

The ego network of DRFZ has 26 direct connections, 15,38% of these are located in the Berlin region and all have connections outside of the region.

Eberhard Karls Universität Tübingen

In the ego network of the Eberhard Karls university are 285 organizations, 3,5% of these are located in the region of Tübingen, all of these are participants of the German subsidy program. 7 out of 10 organizations only have local connections, mostly to each other. In many of the projects (which mostly have only a few participants) these organizations participate in, the Eberhard Karls university is the only organization with connections outside the region. This might be a genuine geographical gatekeeper, following the many economic geographic studies discussed in the theoretical chapter. Primary data should be collected to verify the role of gatekeeper of the Eberhard Karls university in this region.

Archimedes Foundation

The Archimedes foundation is the only foreign gatekeeper in the combined network. Its ego network consists of 16 actors, none of which are located in Talinn. Because this study only includes connections of projects with German organizations, data may have been lost concerning the connections of the foreign actors. Therefore, no conclusions on the foreign actors can be drawn.

Steinbeis Forschungs- und Entwicklungszenteren GmbH

The region of Stuttgart has 7 organizations in the combined network. Among the gatekeepers, only Steinbeis is located there. In the ego network of Steinbeis are 45 organizations, 0% of these are located in the Stuttgart region.

Medizinische Hochschule Hannover

The Hannover region also has 7 organizations in the network, and only this gatekeeper is located there. In the ego network of Hannover medical school are 256 organizations, 1,56% of these are located in Hannover, and one of these organizations (Phenos GmbH) is only connected to other regions through the Hannover medical school. For this firm the school is a geographical gatekeeper.

This combined network does not show evidence of the relevance of geographical proximity in this innovation network. The gatekeepers that are found do not operate on a geographical scale, with the exception of one, and even this is a very small part of the connections of this organization. Because the method for this study is quite crude, real conclusions cannot be drawn with regard to the role of geography in innovation networks, but no evidence has been found implying that organizations in a region have to rely on a single actor for inter-regional connections, at least not in the biotechnology industry.

netdraw visualizations

Netdraw (Borgatti, 2002) is the network visualization software that is used along with UCINET. It helps to depict the networks as a series of lines and dots that represent the actors and their connections. Most networks in this study are too large to be effectively represented by Netdraw, even most ego networks of the gatekeepers are too large to distinguish. There are however some networks details in this study that can be effectively depicted with Netdraw visualizations.

Figure 4.1 shows the relations of the 9 gatekeepers amongst one another.

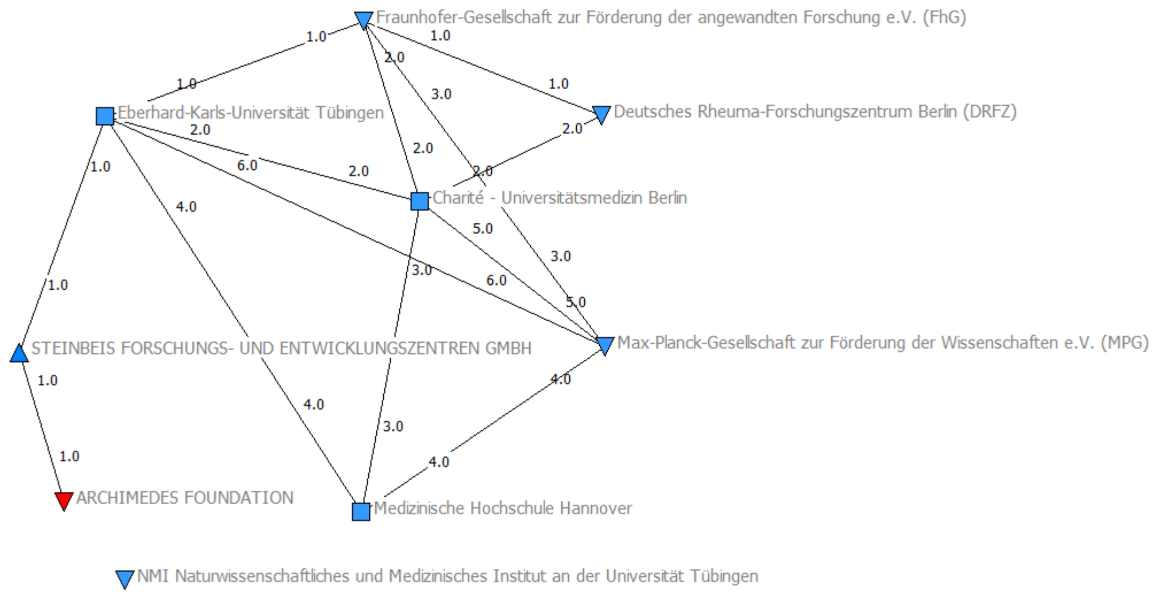


Figure 4.1 links between gatekeepers in the combined network

The triangles are research organizations, the squares are universities. The upward triangle is the private organization and the red triangle is the non-German one. The numbers along the links are the number of times the organizations are linked to each other through various projects, showing the weight of the connections. This diagram shows that the gatekeepers, with the exception of the NMI institute (also figure 4.2), are pretty well connected amongst themselves. The diagram depicts quite effectively the order of the list of gatekeepers, as the Charité is the best connected among the most 'powerful' of the combined network, with only NMI as an exception.

Because the number of direct connections is clearly not the reason for the high listing of NMI on the list of gatekeepers, its ego network is also the one that can be visualized most clearly. Figure 4.2 shows a example of the ego network of one of the gatekeepers.

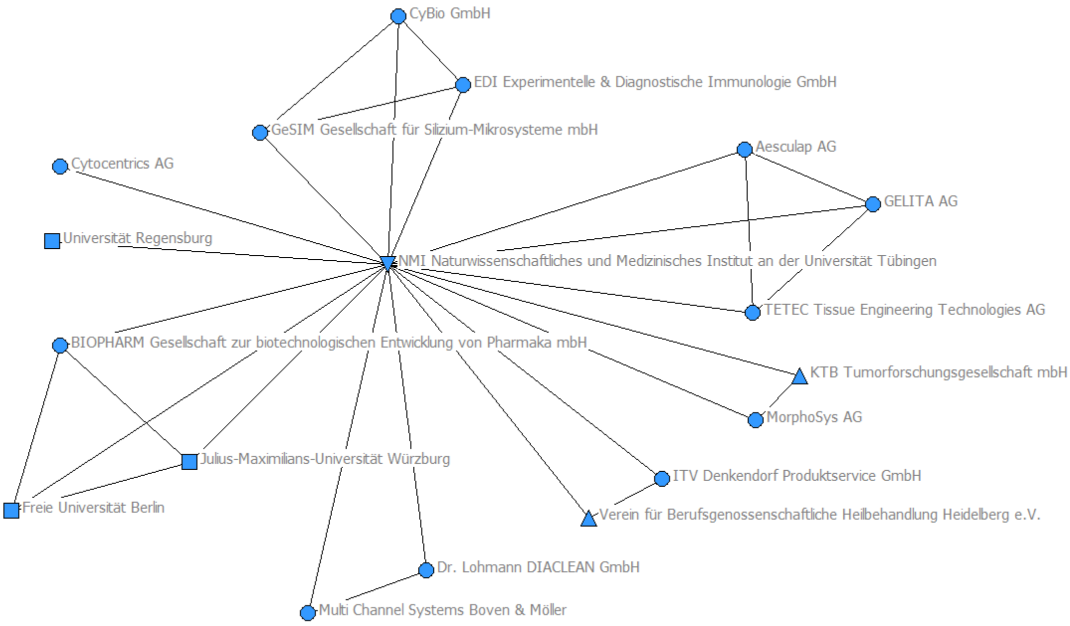


Figure 4.2: ego network (+1) from NMI institute

Figure 4.2 displays the ego network of the organization that has no direct links to any of the other gatekeepers. It depicts how the research institute connects a wide array of different organizations. It also visualizes the low connectivity of the different firms involved, and how research organizations tend to play a connecting and gatekeeping role for the firms. These organizations could have their own links to other actors that are not linked to NMI as well, which are not displayed here. The Free university of Berlin for example, was a gatekeeper in the GER network itself. Figure 4.2 nonetheless gives an indication of how a gatekeeper is a central player in a web of actors.

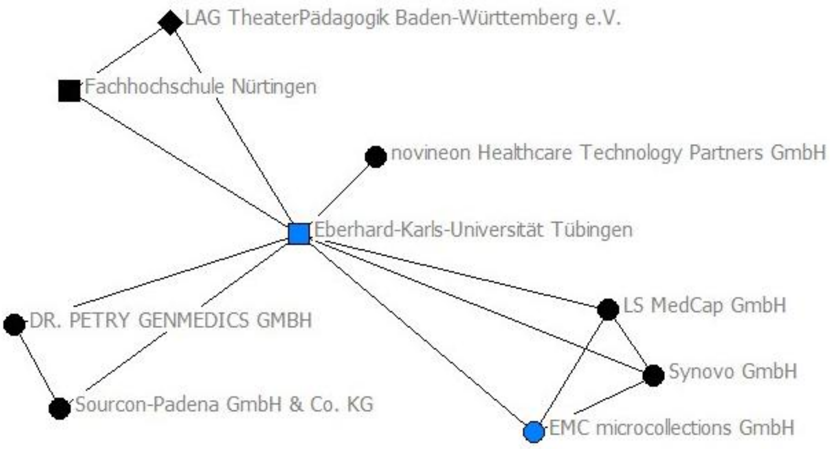


Figure 4.3: geographical gatekeeping network for Eberhard Karls university

Because the Eberhard Karls university proved to be a potential geographical gatekeeper, a Netdraw visual is depicted for the ego networks of the geographically proximate and cognitively bounded nodes around this gatekeeper. These organizations are only connected inter-regionally through either the Eberhard Karls university or EMC Microcollections (Figure 4.3). The black nodes are cognitively dependent on the blue nodes, all nodes are physically located in the Tübingen/Reutlingen county. The circles are firms, the diamond is a private non-profit foundation and the boxes are education institutes. As explained above, these cognitive links are only based on the subsidy data from the two programs researched in this study.

The combination of the networks could have brought a completely new network structure, but joining and comparing the two networks has not yielded considerable surprises. The list of gatekeepers is a combination of the two program lists, the characteristics are more or less the averages between the two subsidy networks. The combination does offer a network model of collaborating organizations that resembles the real world more than the separate networks, because it displays more existing links between actors. There are however, still problems with the way the network is built up. This will be addressed in the discussion chapter. The next paragraph delves deeper into the actual organizations that have been determined gatekeepers, in order to provide some background to the mathematical puzzle that was this chapter.

4.5. qualitative approach

In this paragraph I discuss the third research question.

> *3. Are the results generally applicable in innovation networks of emerging technology markets and what are possible policy implications?*

On a technical level, the combined network that was constructed in the previous paragraphs does resemble the European subsidy network for biotechnology to a greater extent than the smaller networks of the separate subsidy programs did. The conditions in which gatekeepers are important in networks have been described for a large population of actors from different programs. Because they largely displayed similar characteristics, it could be proposed that the technical results may be applicable in other innovation networks in high-tech industries as well.

Governments play a large part in collaborative subsidy networks, because they set the rules for the programs. That is why it is useful to learn about the organizations that have powerful positions in the networks that governments have created in the present programs. In this paragraph, I describe a selection of gatekeepers from the combined network in a concise manner, focusing on the more qualitative aspects of the organizations. Their annual reports, mission statements and other freely available information will be used for this analysis in combination with what I already have learned about them in the previous chapters. This may improve understanding their role as gatekeeper in the subsidy programs.

Four organizations are chosen to be described here. I chose these because they are organizations of different types and sizes, so they cover a wide array of gatekeepers. The Archimedes foundation is not included here, because it is hard to determine if its position on the network would be the same if the data was not focused on Germany. There is a chance that a lot of the direct connections of the Archimedes foundation are not present, and that this is the reason it technically emerges as a gatekeeper for the organizations it connects to the rest of the network. The four organizations that are analyzed are:

1. Charite – University hospital
2. Max Planck – Independent public research society
3. Eberhard Karls Universität - University
4. Steinbeis – 'Private' research center

The gatekeepers are described according to retrieved information from freely accessible sources like their official websites, their annual reports and catalog websites. The budget numbers are the most recent that are found, instead of the figures from the research period, which were not accessible for most organizations. The figures for 2011 still provide a good indication of the organizations income. Finding an independent benchmark that makes it possible to qualitatively compare these organizations proved impossible, as the organizations are very different in kind and scope. Therefore, these summaries are drawn from annual reports, which have a legal obligation to reflect the truth. Of course, in the case of mission statements, this is harder to verify than with employee and budget numbers.

1. Charité medical university Berlin

With 12,908 employees, the Charité medical university is one of the largest university hospitals in Europe, partly because it is the merged medical institute of the Humboldt university and the free university of Berlin, both also gatekeepers in the German subsidy program network. This may help to explain its importance to the cohesiveness to the combined network. Charité is located in Berlin, the largest cluster for the biotechnological sector, it is a public organization and it internally combines the network connections of two large universities. In the study of the ego network of Charité, it is noteworthy to mention that most of the collaborating organization were also large and well connected. The mission statement of Charité is : “We bring the forefront of medicine to the hospital bed”. This means that the organization tries to combine new research with practical treatment, which is why serious, rare and complex illnesses are the specialty of the hospital (Charité website, 2012). In 2011 the budget of Charité was €1,3 billion, from which €158 million was private and the rest public funding (Charité annual report, 2012). In the FP6 and German programs, in the timeframe of this research, Charité helped collecting €80,022,311 in funding for the projects.

Charité medical university participates in 31 projects, 20 projects in the German subsidy program, and 11 in FP6.

2. Max Planck society

The Max Planck Society for the Advancement of Science (MPG) is an independent but non commercial organization. With nearly 80 research institutes under its wings, it has by far the biggest share in both of the datasets of this research. MPG is registered as an association in Berlin, but its headquarters is in München. Many of the research institutes are located elsewhere in Germany, which makes it hard to pinpoint and utilize geographical properties to this organization. The mission of MPG is to promote basic research at its own institutes. In 2011 MPG had 17,019 employees and its budget was €1,9 billion. €1,46 billion came from public sources, the rest was acquired at private benefactors (Max Planck annual report, 2012). In the timeframe of this research, MPG helped acquiring €169,154,798 for projects in the FP6 and GER programs.

The Max Planck Society is present in 65 projects, 45 projects in the German subsidy program, and 20 in FP6.

3. Eberhard Karls Universität

The Eberhard Karls university in Tübingen is one of the oldest and most classical universities in Germany. As one of many universities in our research, it is a large organization with more than 10,000 employees. Eberhard Karls is and wants to stay a world player in research, this is also mentioned in their mission statement. Universities often are the gatekeepers in the networks of this study. As large, well funded organizations full of highly intelligent people without a commercial goal, that specialize in acquiring funding for research, the potential as connector of other organizations is imaginable. In 2011, the university of Tübingen had a research budget of €119,6 million, €79,3 million of which came from public sources and €40,3 from private ones. The combined projects in which the university took part that were studied in this research got €71,423,906 in funding.

The Eberhard Karls university is present in 26 projects, 15 projects in the German subsidy program, and 11 in FP6.

4. Steinbeis research center (GMBH)

As part of a university, the Steinbeis research center looks and feels entirely like a public organization. Since 1997 the technology transfer department is entirely placed under a non-for-profit company, after a series of real estate acquisitions which proved to be unlawful after a tax policy change ('Steinbeis 1983-2008', 2011). Therefore, technically the research organization is a commercial and private entity. The goals of this organizations are promoting the overall economy, and the center is allowed to pursue developments for commercial use. This is also incorporated into the mission statement, which comes down to 'bringing academic research to the market'. The Steinbeis center has 1,462 employees and had a budget of €143 million in 2011. Unfortunately, a division of public and private funds is not available. The funding for the projects in FP6 and GER in which Steinbeis took part is a combined €10,608,814.

The Steinbeis research center participates in 4 projects, all in the FP6 subsidy program.

The organizations that turned out to be gatekeepers are hard to describe by their general, publicly available characteristics. The importance of the public mission was already pointed out in the data research, and the vast differences in employees and budget only prove that these do not unambiguously signify the traits of a gatekeeper. Also lacking is an independent, un-biased source of data on these actors, in which the same data in the same formats can be found on, for example, the budgets or the innovative performance. However, these qualitative properties might be useful in combination with what was already learned in the quantitative part of the study.

Unfortunately, clear policy implications cannot be derived from these findings. It seems that only a more thorough study based on primary data will yield clear conclusions about what exactly makes these organizations the gatekeepers of knowledge in this network based on qualitative charactersitics.

4.6. final remarks

The results from paragraph 4.1 through 4.4 may help to answer question 3 to some extent, because some characteristics of the gatekeeper have been pointed out by the datasets. Universities and research organizations usually hold very important positions in knowledge networks. These types organizations are almost always present in collaborative subsidy projects, and form the links to most other organizations in the network, especially when they are large themselves.

The gatekeepers seem to be large organizations, not necessarily with a large amount of direct links to other organizations, but with a widespread reach in the academic and economic environment in which they operate. Research organizations and universities have contributing characteristics that let them operate on an influential network level. To name a few, highly educated employees, a position in the world that lies between academia and business, worldwide connections with similar organizations without hazards of direct commercial competition, and usually shared goals and good connections to governments.

Further answering of the third research question will be handled in the chapter 5: conclusions.

5. conclusions

Starting this research, I set out to find specific network conditions that stimulate innovation. Governments, universities, research organizations and firms can learn from these conditions, in order to strengthen the position of their interest concerning innovation. The following research questions were posed, in relation to the importance of organizations' postitions in innovation networks:

1. *Which organizations can be identified as gatekeepers in the innovation networks of German led projects within the EU Framework Programme 6 and projects within the German subsidy program between 2003 and 2006 in the biotechnology sector? What are their characteristics? To what extent does the geographic position of these gatekeepers contribute to their role within the network?*

2. *If any, what are the similarities and differences between the characteristics of the gatekeepers in the EU Framework Programme 6, the German subsidy program and the combined network of both programs?*

3. *Are the results generally applicable in innovation networks of emerging technology markets and what are possible policy implications?*

In this study, the characteristics of German subsidy networks in the biotechnology industry are described. These characteristics show what the positions of organizations in the networks are, how these positions contribute to the transfer of knowledge, and what the specific characteristics of gatekeeper organizations in these networks are. First, the findings of this research on gatekeepers in the investigated networks are summarized and compared with the theoretical chapter. Then, the meaning and implications of these results are discussed.

gatekeepers

The gatekeepers in both subsidy networks and in the combined network are all organizations with a public mission, meaning that they are either research organizations or universities. Acting as connectors, public organizations improve the knowledge transfer in a network, while firms with a commercial mission practically apply innovations for commercial success. In the combined network, there are slightly more research organizations than universities acting as gatekeepers, while there are large differences in the separate subsidy programs. From this study I cannot conclude that one or the other organization type is better suited for the role of gatekeeper.

This conclusion is in line with the studies that were discussed in the theoretical chapter. According to Kauffeld-Monz & Fritsch (2008), gatekeepers are likely to be R&D organizations and universities. The latter more so than the former, because of the international relationships universities tend to have, more than most research organizations. This is also supported by Graf (2011), who argued that organizations with a public mission are important to the cohesiveness of a network in the role of gatekeepers, and that their goal should be focused on the improvement of external relations. This is in fact also in line with Allen's (1977) definition of the technological gatekeeper, which stresses that a gatekeeper is an independent, highly connected expert in the technological field.

Generally, the gatekeepers are organizations that are among the most active in the networks, participating in many different innovation projects. I cannot conclude however, that this is a mandatory trait of a gatekeeper: some gatekeepers only have a few connections. These organizations are usually smaller in terms of employees and operating budget, and may be gatekeepers because of their strategic position in the network, connecting parts that are unconnected otherwise. The well connected gatekeepers usually have large numbers of employees and a large budget, such as the Max Planck organization. These may be gatekeepers simply because the network would not be very well connected without them. Small or large, both types of gatekeepers fit the definition of a gatekeeper according to this study well.

The gatekeepers are all located in the established biotechnology clusters in Germany. However, the gatekeepers are primarily connected to organizations outside of the cluster. A study into the localized nature of connections shows that there is only one gatekeeper that seems to play a gatekeeping role on a geographical dimension. Although a more comprehensive study is needed to be able to conclude this, it seems that the gatekeepers do not perform a geographical role for the actors in this network. Moreover, geographic proximity does not seem to be a factor in the distribution of links in this network, as there is a very small chance that linked actors are located in the same county. However, there may well be a lot of actors that are not part of this network that depend on the network organizations acting as geographical gatekeepers.

combining the two programs

There are differences in the two separate subsidy programs, which means that combining them provides a more representative network. The Sixth Framework Programme has larger project groups and is therefore more densely connected. The German program has the opposite characteristics. There are also differences in the distribution and network roles of types of organizations between both subsidy programs. In combining them, a broader image of subsidy networks is presented, resembling a more representative innovation network.

FP6 shows that the international program has more research organizations in a gatekeeping role, while the German program show a majority of gatekeepers that are universities. This is not as expected, in the theory of this study I assumed universities tend to operate on an international stage more often than research organizations.

practical implications

The role of geographic proximity does not seem to be an important factor in these cognitive networks. High-tech organizations can easily find their collaborators outside their physical location without having to rely on a gatekeeper. However, the way the network is organized, gatekeepers still exist on a cognitive level.

The role of geography was more or less assumed in most gatekeeper studies that inspired this research, by setting the point of view inside a geographically bounded area like a cluster of an industrial district. My study implies that, for the biotechnological industry, geographical distance between organizations does not act as a boundary for collaboration, this is at least true in the subsidy networks analyzed here. The geographical point of view thus needs to be assessed more critically. Further studies should investigate if this finding is an anomaly, if and how the importance of physical proximity changes through time and if this is also true for other industries. I elaborate on this in the discussion chapter.

These findings have some implications for the results of gatekeeper studies. For example, the effect of lock-in in clusters is often used as an important reason for the existence of a gatekeeper in a physically bounded area. This is for example an important factor in government policy concerning stimulating intra regional and international connections for its local of regional business milieu. If lock-in does not take place on a geographical level, but on a cognitive network level, the prevention of lock-in would need a completely different approach, with more specific knowledge on the connections in a network.

Gatekeepers are public organizations: The universities and research organizations are the gatekeepers of these innovation networks. The collecting and spreading of knowledge is inherent in the missions of these organizations. Governments that want to stimulate the cohesiveness of innovation networks might want to assess the efficiency of universities and research organizations, probably especially towards the inclusion of firms in projects. Attracting the right collaborators seems to be the most important feature of these organizations.

Because it seems that it is not the number of connections that is important, the quality of these connections might prove to be an interesting research object. For this, primary data research is needed into the connections of gatekeeper organizations. Especially the exact characteristics of the connections of the gatekeepers with a small amount of ties could show the details of the network importance of these gatekeepers.

The gatekeeper as a player in knowledge networks is a very interesting research object, because its existence, its characteristics and its position in a network can describe, assess and determine the cohesiveness and effectiveness of a complete network. The detailed study of these players on the micro level can help improve the effectiveness of organizations, while gatekeeper research on the macro level can help assess and improve global knowledge networks. Altogether, research of gatekeepers can therefore help understanding knowledge networks and the process of innovation. My research is but a small step in this process, but hopefully, by suggesting new questions and a different direction, it is a useful one.

Laurens Hulshof
August 2012

6. discussion

In this chapter I would like to offer some reflection on this study, discussing the overall research as well as following the critical notes made in the text that refer to this chapter. Additionally, I would like to make some suggestions for further research on the subject of innovation networks and gatekeepers.

The origin of this research was when I acquired the very interesting datasets on the Sixth Framework Programme (FP6) and the German subsidy program. Both datasets had been utilized in separate but similar studies, but they had never been matched and joined in order for research on a combined network. This was the premise of my study when I first started working on this thesis.

When the data was matched and completed for the goals I had set for this research, the timeframe of the data was limited to the running time of FP6. This meant that a network analysis on this data would be one with a static perspective, which is unfortunate considering the emphasis on the need for research with an evolutionary perspective in economic geographic network studies (see for example Glückler, 2007; Boschma & Frenken, 2010). However, because this study is carried out in quite an unconventional manner (by using subsidy data, cognitive distance and a broad definition of a gatekeeper), the static nature of the data made the overall process less complicated. If further studies are performed in a similar way, I would recommend adding an evolutionary viewpoint. Because I assume innovation networks in high tech industries are quite versatile (as argued by Gay & Dousset, 2005 and Boschma et al., 2011), studying the way they change over time would be an important addition towards the goal of understanding knowledge flows in the innovation process.

In order of getting to know more about the importance of geographic proximity in these networks, the ego networks of all actors in the network should be investigated for geographic location. In this study, only the gatekeepers are extensively investigated. This limited research already showed that almost all actors that are linked directly to the gatekeepers, also had other direct connections with actors outside of their geographic area. This in itself means that these gatekeepers are not geographic entities for this network, but it is possible that, for example, the *geographic* gatekeepers are different actors altogether. In order of finding this out, the ego networks of all actors need to be investigated in relation to their locations. Because geographic patterns are not the main objective of this study, and investigating all ego networks would be a very time consuming routine, this part is left out in this research.

For this research I limited myself to the secondary data to which I had access. I only ventured into detailed study of individual actors when I had already calculated a list of gatekeepers on a technical basis, from the secondary data. On those gatekeepers I performed some additional network analysis and a limited, web based investigation. The most important critique on this study is that a connection from a secondary source does not prove a robust relationship between organizations. Without information from primary sources, be it from interviews, surveys or direct company records, the links between the actors in these networks cannot be assumed as structural knowledge transfer.

In this study, it is assumed that knowledge transfer occurs when two organizations have collaborated in an R&D project for which they received funding. The outcome of the projects are to be used by all the parties equally, but this does not mean a robust relationship has developed. The possibility of knowledge transfer is certainly high between subsidy project collaborators, as they work together on a single project of innovation, and are obligated to share all results, but there is no guarantee of organizations actively learning from each other. In this light, the connections between actors in this study can be seen as indications of knowledge networks, but not as irrefutably existing knowledge flows.

As explained in the theoretical chapter, hard proof of actual knowledge transfer is almost never acquired in this field of study, but surveys and interviews can be used to confirm actual meaningful relationships between organizations, as these often include a human aspect. Also mentioned, primary data is also too time consuming to acquire for a large scale network. A smaller scaled study with primary sourced data could be based on networks surrounding the gatekeepers of this study, in order to verify some of the conclusions of this study.

This study concludes that the size of organizations does not necessarily facilitates gatekeeperness, but that a majority of the gatekeepers are (very) large operations. Larger organizations have a higher chance of being a gatekeeper, but in a larger organization the chance is also higher that two people within the organization are not in contact with each other. The Max Planck society for example, consists of many departments located all over Germany, which could an explanation of it being a gatekeeper, but also raises questions on the value of an outside connection of one of its departments to a coworker in another department. Suspicion of a large organization could be ungrounded, because the knowledge transfer within an organization could be very good, but this also has to be determined by acquiring information from a primary source.

Based on these points and on the conclusions of my research, I would like to suggest some further study on the following subjects.

Adding primary data to smaller groups of network actors of the present study would be valuable. The networks surrounding the identified gatekeepers should be investigated more closely, especially the gatekeepers themselves. A survey conducted in a larger group of actors would be practical for getting to know actual knowledge flows and their intensity. Interviews with key players in the gatekeeper organizations would be a useful method for uncovering the true innovation process, because in depth workings of collaborations is probably a human-centered and variable process for every different organization.

In order of finding out the actual innovative performance of innovation networks, patent records could be combined with the subsidy records and findings of this study. The present study maps out the knowledge networks, but it is not possible to indicate which combinations of actors, network characteristics or other attributes contribute to actual innovation. As explained in the methodological chapter, patent records do not quantify innovative success completely, but they could form a useful addition to the present study.

As mentioned above, the study could be expanded in a useful way by performing the same actions in different timeframes, preferably with the exact same networks. A lot of variation in actors and attributes is expected, making an evolutionary study of the network buildup alone very interesting. Which actors are the same, and how does their position in the network change? A dynamic study on innovation networks will inherently have a high level of complexity, due to different scale levels and different timeframes (Gay & Dousset, 2005; Katz, 2006), but can help understanding the innovation process by investigating how volatile a snapshot like the present study actually is.

Another suggestion for further study is recreating this research in different high tech industries, in order to find out if the biotechnological industry is representative for other emerging markets. Conclusions could then be used in a broader perspective, which makes them more useful in a practical setting, for future industries as well as existing ones.

A C K N O W L E D G E M E N T S

Andrea Morrison and Tom Broekel, thank you for providing the academic support during the writing process.

Ruth, a few words on this page can never do justice to the significance of having you by my side. Nevertheless I thank you for your support and advice during the writing, evaluating, rewriting and rewriting, but most importantly for your love and your bad jokes. Thank you for being and living with me.

“Oké het duurt verrèkte lang, maar ooit siert een diploma dit behang…” In desperate moments these words, framed on the wall next to my computer screen, have reminded me of the unlimited support from my father and mother. Hans, Annie, thank you for the preposterous amount of patience and confidence.

Joost, Vivien, Martine and Erik: Thank you for your boosts, cheerups, good times and advice (coming mostly from first-hand experience). Can’t beat a family like this.

Olief, thank you for joining your big sister. You two cheer everyone up as a team now!

L I T E R A T U R E

Allen, T. J. (1977), Managing the Flows of Technology: Technology Transfer and the Disseminatin of Technological Information within the R&D Organization. *MIT Press.*

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

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

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

Boschma R. A., Iammarino S (2007), Related variety and regional growth in Italy: working paper no. 162, SPRU. University of Sussex, Brighton

Boschma, R.A.,Ter Wal, A.L.J. (2007), Knowledge networks and innovative performance in an industrial district: The case of a footwear district in the south of Italy. *Industry &Innovation* 14 (2), 177-199

Boschma, R.A., Frenken, K. (2010), The spatial evolution of innovation networks. A proximity perspective. in R.A. Boschma, R. Martin (eds.), Handbook on Evolutionary Economic Geography. Cheltenham: Edward Elgar, 120–135.

Borgatti, S.P., Everett, M.G. and Freeman, L.C. (2002), Ucinet for Windows: Software for Social Network Analysis. Harvard, MA: Analytic Technologies.

Borgatti, S.P., (2002), Netdraw Network Visualization. Analytic Technologies: Harvard, M.A.

Borgatti, S.P., (2005), Centrality and network flow. *Social Networks* 27 (1), 55–71

Borgatti, S.P., Everett, M.G. (2006), A Graph-theoretic perspective on centrality. Social Networks 28, 466–484

Braczyk,H., Cooke, P., Heidenreich, M. (eds.) (1998), *Regional Innovation Systems.* UCL Press: London.

Breschi, S., Lissoni, F. (2001), Knowledge spillovers and Local Innovation Systems: A critical Survey

Breschi, S., Lissoni, F. (2003), Mobility and social networks: Localised knowledge spillovers revisited. *CESPRI Working Paper* 142, Bocconi University, Milan.

Broekel, T., Hartog, M. (2011), Explaining the structure of inter-organizational networks using exponential random graph models: does proximity matter? *Papers in Evolutionary Economic Geography* # 11.07

Camagni, R. (1991), *Innovation networks: Spatial perspectives*. London: Belhaven Press.

Carlsson, B. (2006), Internationalization of innovation systems: A survey of the literature. *Research policy* 35 (1), 56-67

Castells, M. (1997), An introduction to the information age, *City*, 2 (7), 6-16

Cooke, P. (2001), Regional innovation systems, clusters, and the knowledge economy. *Industrial Corporate Change* 10, 945–974

Edquist, C. (1997), Systems of innovation approaches—Their emergence and characteristics. In *Systems of innovation—Technologies, institutions and organizations*, edited by Charles Edquist, 1-40. London: Pinter.

Freeman, C. (1991), Networks of innovators: A synthesis of research issues, *Research Policy,* 20, 499-514

Freeman, C. (1995), The 'National System of Innovation' in historical perspective. *Cambridge Journal of Economics*. 19, 5-24

Freeman, L.C., (1979), Centrality in networks: I. conceptual clarification. Social Networks 1, 215–239. Connoted with initials in text

Freeman, L.C., Borgatti, S.P., White, D.R. (1991), Centrality in valued graphs: a measure of betweenness based on network flow. Social Networks 13, 141–154. Connoted with initials in text

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

Fritsch, M. (2002), Measuring the Quality of Regional Innovation Systems: A Knowledge Production Function Approach. *International Regional Science Review* 25 (1), 86–101

Fritsch, M., Graf, H. (2010), How General Conditions Affect Regional Innovation Systems – The Case of the Two Germanys, *Jena Economic Research Papers* 2010 – 054

Giuliani, E., Bell, M. (2005), The micro-determinants of meso-level learning and innovation: evidence from a Chilean wine cluster. *Research Policy* 34, 47–68

Giuliani, E. (2007), The selective nature of knowledge networks in clusters: evidence from the wine industry, *Journal of Economic Geography* 7(2), 139-168.

Glückler, J. (2007), Economic geography and the evolution of networks. *Journal of Economic Geography* 7, 619–634

Grabher, G., Ibert, O. (2006), Bad company? The ambiguity of personal knowledge networks. *Journal of Economic Geography* 6, 251–271

Graf, H. (2006), The network of innovators in Jena: An application of social network analysis. *Research Policy* 35 , 463–480

Graf, H. (2011), Gatekeepers in regional networks of Innovators. *Cambridge Journal of Economics* 2011 35, 173–198

Hagedoorn, J. (1993), Understanding the rationale of strategic technology partnering: Interorganizational modes of cooperation and sectoral differences. *Strategic Management Journal* 14 (5), 371–85

Hanneman, Robert A. and Mark Riddle (2005), Introduction to social network methods. *Riverside, CA*: University of California, Riverside

Haythornethwaite, K. (1996), Social network analysis: An approach and technique for the study of information exchange, *Library & Information Science Research,* 18 (4), 323–342

Jaffe, A. B., M. Trajtenberg, et al. (1993), Geographic localization of knowledge spillovers as evidenced by patent citations. *Quarterly Journal of Economics* 63 (3), 576-598

Katz, J. S. (2006), Indicators for complex innovation systems. *Research Policy* 35, 893–909

Kauffeld-Monz & Fritsch (2008), Who are the brokers of knowledge in regional systems of innovation? A multi-actor network analysis, *Jena Economic Research Papers* 2008 - 089

Krugman (1991), Increasing Returns and Economic Geography. *Journal of Political Economy, University of Chicago Press*, vol. 99(3), 483-99

Lambooy J. G. and Boschma R. A. (2001), Evolutionary economics and regional policy. *Annals of Regional Science* 35, 113–131

Lehrer, M. (2005), Science-driven vs. market-pioneering high tech: comparative German technology sectors in the late nineteenth and late twentieth centuries. *Industrial and Corporate Change* 14 (2), 251-278

Lundvall, B.Å. (1985), Product Innovation and User-producer Interaction, Industrial Development. *Research Series 31,* Aalborg: Aalborg University Press

Maggioni, M.A., Nosvelli, M., Uberti, T.E. (2007), Space versus networks in the geography of innovation: a European analysis. *Papers in Regional Studies* 86 (3), 471-493

Maggioni, M.A., Uberti, T.E. (2009), Knowledge networks across Europe: which distance matters? *Annals of Regional Science* 43, 691–720

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

Morgan, K. (2004), The exaggerated death of geography: learning, proximity and territorial innovation systems. *Journal of Economic Geography* 4, 3-21

Morrison A (2008), Gatekeepers of knowledge within industrial districts: who they are, how they interact. *Regional Studies* 42, 817–835

Morrison, A. Rabelotti, R., Zirulia, L. (2012), When do global pipelines enhance knowledge diffusion in clusters?, forthcoming in *Economic Geography*

Nooteboom, B. (2003), Problems and Solutions in Knowledge Transfer. *Paper for the conference on "The influence of co-operation, networks and institutions on regional innovation systems"* Max Planck Institute, Jena, 8-10 February 2001.

Ozman, M. (2009), Inter-firm networks and innovation: a survey of literature *Economics of Innovation and New Technology* 18 (1), 39-67

Ponds, R., Van Oort, F.G., Frenken, K. (2007), The geographical and institutional proximity of research collaboration. Regional Science 86, 423–443

Porter, M.E. (1998), *On Competiton,* Harvard Business School Press: Cambridge, M.A.

Powell, W. W. (1990), Neither Market Nor Hierarchy: Network Forms of Organization. *Research in Organizational Behavior* 12, 295-336

Powell, W.W., Koput, K.W., Smith-Doerr, L. (1996), Interorganizational Collaboration and the Locus of Innovation: Networks of Learning in Biotechnology. *Administrative Science Quarterly* 41 (1), 116-145

Sabidussi, G. (1966), The centrality index of a graph. *Psychometrika* 31, 581–603.

Singh, J. (2005), Collaborative networks as determinants of knowledge diffusion patterns, *Management Science* 51 (5), 756-770

Sixth framework programme practical guide (2003), found online in october 2011: http://europa.eu.int/comm/research/fp6/index_en.html

Ter Wal, A.L.J. & R.A. Boschma (2009), Applying social network analysis in economic geography: framing some key analytic issues. Annals of Regional Science 43 (3), pp. 739-756

Ter Wal, A.L.J. (2010), The dynamics of inventor networks in biotechnology: geographical proximity versus triadic closure. Best Paper Proceedings of the 2010 Academy of Management Meeting

UCINET 6, Borgatti, S.P., Everett, M.G. and Freeman, L.C. (2002), Ucinet for Windows: Software for Social Network Analysis. Harvard, MA: Analytic Technologies.

Uzzi, B. (1996), The sources and consequences of embeddedness for the economic performance of organizations: the network effect, *American Sociological Review*, 61(4), 674-698

Von Hippel, E. (1994), Sticky information and the locus of problem solving: implications for innovation, *Management Science* 40, 429–439

Wasserman, S., Faust, K. (1994), Social network analysis: methods and applications. *Cambridge University Press*, Cambridge

Wellman, B. (1983), Network Analysis: Some Basic Principles. *Sociological Theory*, 1, 155-200

Wink, R. (2008), Gatekeepers and Proximity in Science-driven Sectors in Europe and Asia: The Case of Human Embryonic Stem Cell Research. *Regional Studies* 42, 777-791

Zabala-Iturriagagoitia, J.M., Voigt, P., Gutiérrez-Gracia, A., Jiménez-Sáez, F.(2010), Regional Innovation Systems: How to Assess Performance. *Regional Studies*, 41 (5), 661-672