
Munich Personal RePEc Archive

Modeling interregional research collaborations in German biotechnology using industry directory data: A quantitative social network analysis

Mitze, Timo and Strotebeck, Falk

University of Southern Denmark, RWI, RCEA, Fachhochschule Südwestfalen

1 February 2017

Online at https://mpra.ub.uni-muenchen.de/83392/
MPRA Paper No. 83392, posted 22 Dec 2017 04:32 UTC

Modeling interregional research collaborations in German biotechnology using industry directory data: A quantitative social network analysis

Timo Mitze [1,*] and Falk Strotebeck [2]

[1] Department of Business and Economics, University of Southern Denmark; tmitze@sam.sdu.dk, RWI Essen and RCEA

[2] Fachhochschule Südwestfalen; strotebeck.falk@fh-swf.de

* Correspondence: Alsion 2, DK-6400 Sønderborg; Tel.: +45-6550-1770; Email: tmitze@sam.sdu.dk.

Abstract:

We use industry directory data as a novel source of information to model the strength of interregional research collaboration in German biotechnology. Specifically, we gather data on the number of research collaborations for biotech actors listed in the BIOCOM Year and Address book and aggregate this information to the level of German NUTS3 regions. This allows us to set up a modeling framework that treats individual regions as nodes of the biotech research network. We then specify the collaboration activity between regional nodes as a function of research and economic capacities at the regional level, the geographical proximity between regions, and policy variables. Our results show that the strength of interregional research collaboration can be related to both node properties and the relationship between nodes. As such, we find that modern locational factors are positively correlated with the extent of interregional research collaboration, while geographical distance is found to be an impediment to collaboration. The results further show that the pursuit of network and cluster policies in the biotech sector, particularly through collaborative R&D funding, is positively related to the strength of the interregional collaboration activity.

JEL: C21, L65, O38, R38

Keywords: Biotechnology, research collaboration, cluster policy, social network analysis, count data

1. Introduction

Economic systems are characterized by mutual interdependencies among their actors, and the emergence of network and collaboration structures marks a crucial channel for knowledge exchange and diffusion in modern economies. In particular technology-intensive industries are prone towards the development of networks and alliances with interrelated actors as a means to external knowledge access (Quintana-Garcia and Benavides-Valasco, 2004). There is by now ample evidence showing that to gain and maintain access to external knowledge constitutes an essential success factor for 1) building up own knowledge stocks, 2) benefiting from effective knowledge transfer and –ultimately– 3) improving the own research and development (R&D) as well as innovation performance (see, e.g., Kesteloot and Veugelers, 1995; Veugelers, 1998; Bathelt et al., 2004; Paci et al., 2014 for evidence at the firm, regional, and national level). Although empirical research on knowledge sharing and network formation is a challenging task given the inherent complexity of the underlying networks (Schweitzer et al., 2009), it may at the same time yield new insights on the drivers of knowledge networks and thus offer guidance for policymakers on how to develop proper policy instruments to foster knowledge creation and diffusion.

Accordingly, in this paper we aim to map the research collaboration structure in the sectoral innovation network of the German biotech industry and analyse the main determinants for link formation between the nodes of the network. Given that the spatial distribution of the biotech industry is typically characterized by the prevalence of distinct urban centers and regional clusters (see Ter Wal, 2014 for Germany as well as Ó hUallancháin and Lee, 2014 for the American biotech industry), we combine the aggregate analysis of the sectoral innovation system with a particular regional perspective. Our motivation for studying the innovation network in biotechnology stems from two facts: Firstly, from a technological and innovation-centered perspective, biotechnology can be seen as a frontier technology for the invention and commercialization of new products and processes in fields such as health care, agriculture, food, and manufacturing (Kang and Park, 2012; Hazir and Autant-Bernard, 2014). From a policy perspective, the biotech industry can further be

regarded as an interesting case study due to the role played by avant-garde network and cluster policies, such as the *BioRegio* contest and its successors, to foster the competitiveness and innovativeness of the entire industry (Dohse, 2007; Engel et al., 2013). These collaboration-based science and technology (S&T) policy instruments have become a common tool of Germany's high-tech strategy 2020 (BMBF, 2010).

For our empirical analysis, we combine information from multiple sources to map and model the strength of research collaborations in German biotechnology. Our key information on research collaboration activity is thereby extracted from a commercial industry directory of German biotech firms and institutions, the BIOCOM Year- and Addressbook, which collects information on firms and institutions in the German biotech sector as well as provides data on their research collaboration activities. The data is then aggregated to the level of NUTS3 regions (*Kreise*), which allows us to set up a modeling framework for the German biotech research collaboration network using NUTS3 regions as individual nodes of the national collaboration network. The aggregated research collaboration data for NUTS regions are then merged with several indicators related to the region's geography, economy, and innovation system. Finally, using indicators from social network analysis combined with an econometric analysis for count data, we then assess the role played by biotech-specific and general regional resource endowments in determining tie formation between regional nodes. We also analyse the importance of proximity between region pairs as a driver for network formation and quantify the role played by biotech-related policy instruments.

Regarding the latter, we particularly focus on public R&D support schemes to support collaborative R&D activities in selected regional networks/cluster initiatives (so-called *BioRegions*). Fostering collaborative R&D activities within and between regions can be seen as an essential intermediate policy target in order to improve the international competitiveness of German biotechnology (see, for instance, RWI et al., 2014 for a discussion on the issue of enhancing research and innovation collaboration activity as a short-term policy target). If we regard this intermediate target as a necessary prerequisite for overall policy effectiveness, our empirical analysis on network formation may

hence be seen as a complementary approach to evaluate the success of R&D-based cluster policies beyond the scope of a direct assessment of "classical" outcome variables, such as the regional patent activity or start-up rates in the biotech industry (see, for instance, Staehler et al., 2006; Engel et al., 2013 for evidence on the German *BioRegio* contest as well as Uyarra and Ramlogan, 2012 for a meta study on the link between cluster policy and innovation).

Foreshadowing some key empirical results, the descriptive statistics of indicators from social network analysis show that the structure of the research collaboration network in German biotechnology is far from being random and features specific sector-region and overall regional characteristics. Although we find that collaboration activities are generally highly localized and geographical distance works as an impediment to tie formation, selected interregional linkages over long distances are found to mark important channels of knowledge flows and can be seen as crucial bridges between local biotechnology sub-networks. This result is in line with earlier empirical contributions assessing the role of local and global networks for knowledge flows and innovative activity in the biotech industry (see, e.g., Gertler and Levitte, 2005, Maskell et al., 2006).

On the basis of a zero-inflated count-data regression approach we further show that research-related resource endowments at the regional level are positively correlated with the extent of collaboration activity between NUTS3 regions. Modern location factors, such as start-up activities in high-tech manufacturing sectors and knowledge-intensive services as well as agglomeration factors, matter for the formation of network ties, while geographic distance –which is often used as an indirect measure for social proximity in social network analysis– is found to be an impediment to interregional research collaborations. Regarding the role of cluster-based R&D policy, particularly with regard to monetary incentives such as the volume of collaborative R&D funding, we find that these policy instruments are positively correlated with the number of research collaborations between regions in the German biotech network.

The remainder of the paper is organized as follows: The next section introduces the underlying data used to measure collaboration activity in the biotech industry and presents some stylized facts of the research collaboration network in the German biotech industry with the help of indicators from social network analysis. Section 3 then develops an econometric modeling framework to estimate the determinants of collaboration activity between NUTS3 regions using model specifications for count data. The section also presents the empirical results of the estimation approach and conducts robustness tests. Section 4 discusses the main implications of our approach with regard to theory, practice, and policy. Section 5 finally concludes the paper.

2. Mapping interregional research collaborations using industry directory data

In this section, we take a closer look at the research collaboration network in the German biotech sector with a focus on data issues. Readers interested in obtaining general information on the state of the sectoral innovation system in German biotechnology are referred to Appendix A. This appendix also provides an overview of the institutional setup of industry-specific policy support schemes, thereby highlighting the role of distinct regional biotechnology initiatives (*BioRegions*). The background information in the appendix may be helpful for uninformed readers to understand the emergence of urban centers and regional clusters in German biotechnology, which motivates our focus on regional entities as nodes of the German research network in biotechnology.

When it comes to the issue of measuring and mapping research collaboration structures within a particular industry, previous contributions to the literature have proposed indicators that predominantly rely on patent data (e.g., patent citations or co-patenting; see, e.g., Balconi et al., 2004; Ma and Lee, 2008) or public funding information (see, e.g., Scherngell and Barber, 2009). In this study, we make use of alternative firm-level information gathered from a commercial industry directory, the BIOCOM Biotechnology Year and Address book, which contains basic information on firms and institutions of the German biotech sector as well as on their (research) collaboration activities. The BIOCOM AG is a specialized information provider for the European biotech industry and

publishes its Year and Address book on an annual basis. The industry directory has become a main source of information for the German biotechnology. Between 2005 and 2015 the BIOCOM AG has also complemented its industry directory with annual surveys of the structure of the German biotechnology, which have been conducted on behalf of the German Federal Ministry for Education and Research and have been used by the OECD for publishing its "Key Biotech Indicators".

To assess the coverage of the BIOCOM industry directory, we show in Table 1 the number of biotech companies and industry-wide employment levels as reported by alternative industry accounts for the year 2004. As the table shows, the reported number of companies covered by the BIOCOM Year and Address book thereby lies between the narrower definition of biotech core companies applied by Ernst & Young (2005) and the slightly wider definition adopted by the German statistical office (Statistisches Bundesamt, 2005). With regard to the overall employment level in biotechnology, the BIOCOM industry directory adopts the widest definition of all three industry accounts. Taken together, the comparison in Table 1 indicates that the BIOCOM data provide a sufficient coverage of the German biotech sector and can be regarded as being representative for the industry.

Table 1. Estimated size of the German biotech industry based on three different industry accounts

	Ernst & Young	Statistisches Bundesamt	BIOCOM AG
No. of (core) companies	380	572	541
No. of employees	9,703	11,958	14,437

Notes: The data cover the sample year 2004 and have been retrieved from the Internet platform http://www.biotechnologie.de. Core companies (category I) are defined as those using modern bioengineering methods for R&D and production. The method applied by Ernst & Young only covers a selected number of core companies; values have been corrected for methodical changes. *Sources*: Ernst & Young (2005); Statistisches Bundesamt (2005); BIOCOM AG (2005).

For the empirical analysis in the remainder parts of this work, we extract data from the BIOCOM Year and Address book 2005 as main source of information. The sample year 2005 has mainly been selected for two reasons: 1) As shown in the stylized facts of the German biotech industry in

Appendix A, the period covering the late 1990s and early 2000s can be seen as the most dynamic expansion period of German biotechnology. The associated building up of capacities and collaborative linkages should thus be reflected in the industry structure for 2005. 2) Given that the *BioRegio* contest, as a blueprint for a broader program family of cluster policies in Germany, was implemented in the period 1997 to 2001, the year 2005 allows us to assess *ex post* the mid-run effects of the policy scheme on the formation of collaborative linkages among regions in the industry and thus allows us to make statements about the role of policy in shaping a sectoral innovation system with a focus on local cluster initiatives. As a robustness check, we will additionally extract information for the sample year 2009 in order to look at variations in the data over time.

In its main function as an industry directory, the BIOCOM Year and Address book provides basic information on listed biotech firms and institutions (including address and contact information, foundation year, and number of employees). What makes the BIOCOM industry data particularly interesting for our study is that they also contain details on the collaboration activity (names of collaboration partners) for each listed biotech actor. We will use this latter information as the key input for our empirical analysis. However, one limitation of the reported collaboration activity is that the actual type of collaboration activity is not precisely specified in the BIOCOM data. Therefore, detailed Internet research has been conducted to exclude pure supplier relationships (objects of utility such as petri dishes or pipettes) and advisory services (such as consultancies). This reduction leads to a set of 575 core companies of the German biotech sector with R&D as the main business activity and a strong focus on R&D collaboration in 2005.

Due to the interdisciplinary nature of the biotech sector (with strong links to pharmacy or textiles and chemicals; see Cooke et al., 2007), in a second step each listed national collaboration partner not included in the set of core companies has been added to the latter set. This results in a data set of 1002 firms and institutions, which will be used throughout the empirical analysis. At this stage, we exclude foreign collaboration partners of German biotech firms given our focus on modeling the determinants of interregional research collaborations within the German biotech collaboration net-

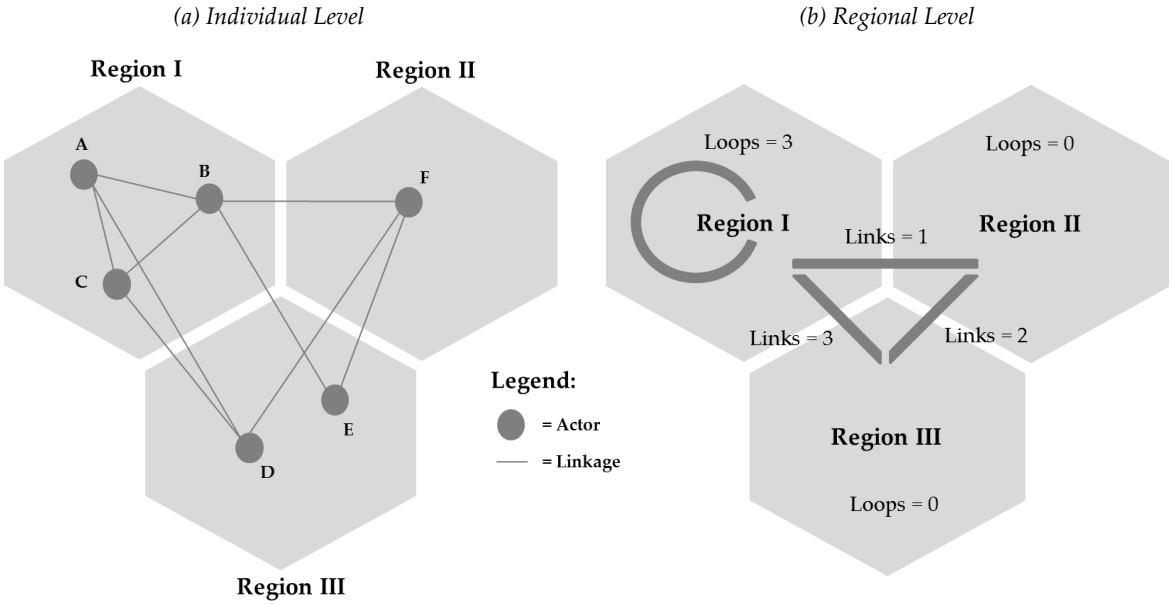
work. For instance, the inclusion of selected international collaboration partners outside Germany may lead to a measurement bias when assessing the role of geographic distance as impediment to collaboration within the German sectoral innovation network in biotechnology. We have thus decided to focus on German regions as a closed geographical system. Clearly, future research should also account for the role of international collaboration activity as outlined in Ma and Lee (2008), for instance.

Starting from the extracted data set of 1002 German biotech actors and their collaboration partners within Germany in 2005, the actor-specific information is then aggregated to the regional level (NUTS3) – thereby reducing the role of individual actors to that of representative agents within a region. For our sample year 2005, 178 out of the 439 German regions are found to host at least one biotech actor listed in the BIOCOM industry directory. On the one hand, the aggregation helps us to cope with the inherent heterogeneity among biotech actors, and on the other hand it further allows us to link data on collaborative linkages between regions with regional covariates in order to characterize the determinants of interregional collaboration activity. Similar covariates are unfortunately not available at the individual actor level. Our method of aggregation is highlighted in Figure 1 for the case of six exemplary biotech actors located in three different regions. While panel (a) in Figure 1 shows the underlying research collaboration network at the level of the individual actors, panel (b) displays the results for the aggregated network at the regional level.

To give an example: Since all three actors (A,B,C) in Region I are engaged in pairwise research collaborations with the other actors, the number of intraregional linkages (loops) in Region I amounts to three. Similarly, since Actors A and C are engaged in collaboration with Actor D in Region III and Actor B has an active collaboration with Actor E in Region III, the total number of collaborative linkages between Region I and Region III is three as indicated in panel (b) of Figure 1. Taken together, aggregating all intra- and interregional linkages to the regional level gives a 3x3 matrix with the number of intraregional collaboration linkages (loops) on the diagonal and interregional linkages in the remaining matrix cells. Since research collaboration activity is undirected, the

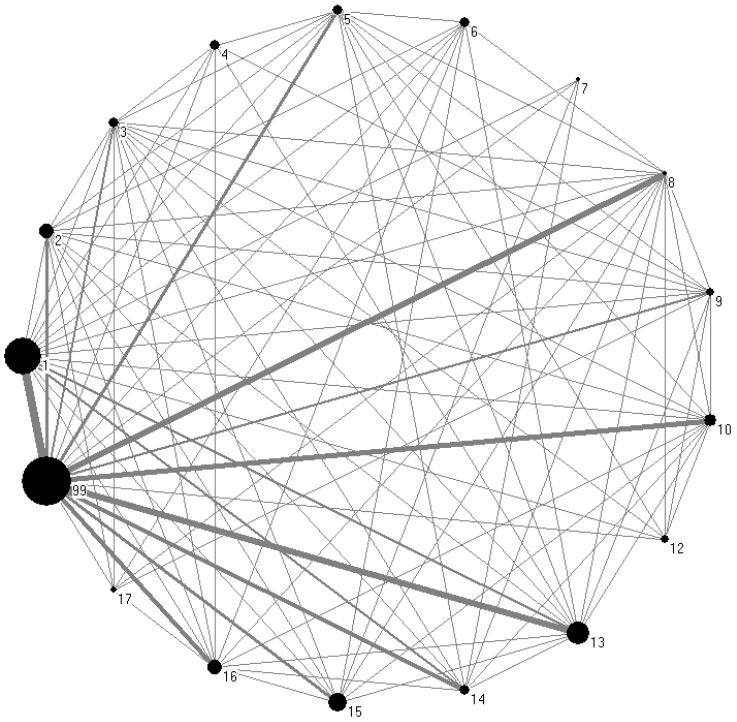
matrix is symmetric. Applied to the case of German NUTS3 regions, this gives a 439x439 matrix measuring the intra- and interregional research collaboration activity in German biotechnology.

Figure 1. Regional aggregation of collaborative linkages among German biotech actors



In a further step, these 178 districts have been mapped into the set of the 17 *BioRegions*, which have been formed in the course of the *BioRegio* contest (see Dohse, 2007 as well as Table B.3 in the appendix). Given that these *BioRegions*, which can be seen as larger macro-regions comprising several NUTS3 regions, are a crucial backbone of the German biotech industry, Figure 2 displays the interregional collaboration network grouped by these *BioRegions*. Each node in Figure 2 presents one *BioRegion*, and ties between nodes visualize collaborative research linkages between the *BioRegions*. The size of the nodes indicates the number of internal linkages (loops) within each *BioRegion*, whereas the width of each tie linking two nodes reflects the number of collaborative research linkages observed between the two macro-regions.

Figure 2. Collaborative linkages within and between German *BioRegions* (in 2005)



No.	Name of *BioRegion*	No.	Name of *BioRegion*
1	BioTOP-Initiative Berlin-Brandenburg	**10**	BioInitiative Nord
2	Region Bremen	**11**	Region Nordwest-Niedersachsen*
3	BioRegion Freiburg	**12**	BioRegion Regensburg
4	BioRegion Greifswald-Rostock	**13**	BioRegion Rheinland
5	BioRegion Halle-Leipzig	**14**	BioRegion Rhein-Main
6	BioRegion Jena	**15**	BioRegion Rhein-Neckar-Dreieck
7	BioMIT Mittelhessen	**16**	BioRegion Stuttgart/Neckar-Alb
8	Initiativkreis Biotechnologie München	**17**	Biotechnologie Ulm
9	BioRegioN	**99**	Not part of established *BioRegion*

Notes: * = not represented in the sample; calculated on the basis of data from BIOCOM (2005), and the definition of *BioRegions* is taken from Dohse (2007).

As Figure 2 shows, the German biotechnology research network is characterized by mutual interdependencies among the *BioRegions*. We also observe a high degree of regional heterogeneity, both with respect to intra- as well as interregional collaboration activity. The rather big "99"-node of regional actors outside any established *BioRegion* demonstrates that German biotech activity is not solely concentrated within these macro-regional cluster initiatives. With 378 actors to be found in the remainder "99"-node, roughly 37 percent of biotech actors from our BIOCOM data set are not located in a particular *Bioregion*; however, this also means that 63 percent indeed are. Among the *Bio-*

Regions the biggest player characterized both with regard to node size and edge width is the BioTOP Initiative Berlin-Brandenburg with 133 actors. As Table 2 summarizes, the remaining *BioRegions* consist of about 30 actors on average.

Based on this basic information, we can compute two further indicators, which have been developed in the field of social network analysis (SNA) to analyse the observed network structure with regard to the position of inidvidual *BioRegions*: 1) the degree centrality (C^D) of a node (n_k) shown in eq.(1) measures the number of direct linkages for underlying actors associated with node n_k and sample size *k*=1,…,N, where the function $f(n_i, n_k)$ counts the number of direct linkages between actors located in nodes n_k and n_i (Freeman, 1978/79) as

$$C^D(n_k) = \sum_{i=1}^{N} f(n_i, n_k). \quad (1)$$

Moreover, 2) the average degree for node n_k can be calculated by dividing $C^D(n_k)$ with the number of actors per node. To give a numerical example: As shown in Table 2, the BioTOP Initiative Berlin-Brandenburg has a total of 133 biotech actors. The *BioRegion* has a total count of $C^D(n_k) = 569$ direct linkages both within the macro-region (denoted as loops for $i=k$) as well as across *Bioregions* (interregional linkages for $i \neq k$). As Table 2 shows the BioTOP Initiative Berlin-Brandenburg has the largest degree centrality as well as the largest average degree centrality. Similarly, the runner-ups BioRegion Rheinland with $C^D(n_k) = 260$ and BioRegion Rhein-NeckarDreieck with $C^D(n_k) = 177$ are also among the *BioRegions* with the largest average degree centrality, which indicates that the number of collaborative linkages seems to be positively correlated with the number of firms and institutions, as indicated in Table 2.

Table 2. Degree centrality and average degree for German *BioRegions*

No. of *BioRegions*	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
$C^D(n_k)$	569	110	78	46	96	41	31	140	55
No. of firms / institutions	133	40	23	18	28	17	11	56	23
Average degree	4.27	2.75	3.39	2.55	3.42	2.41	2.81	2.5	2.39
No. of *BioRegions*	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	
$C^D(n_k)$	142	0	41	260	132	177	156	19	
No. of firms / institutions	47	0	14	74	45	47	40	8	
Average degree	3.02	0	2.92	3.51	2.93	3.76	3.9	2.37	

Notes: Calculated on the basis of data from BIOCOM (2005); *BioRegions* are defined in Figure 2.

However, this relationship does not appear to be a linear one: For example, the BioRegion Freiburg (No. 3) and BioRegioN (No. 9) both contain 23 biotech actors, but there is a big difference with regard to the regions' average degree centrality (2.39 compared to 3.39, respectively). This heterogeneity also becomes visible if we compare the BioRegion Rhein-Main (No. 14) with BioRegion Stuttgart/Neckar-Alb (No. 16): While Rhein-Main contains more actors (firms/-institutions) than Stuttgart/Neckar-Alb (45 compared to 40), the latter has a much higher average degree (2.9 compared to 3.9). This indicates that there is a more complex story to tell rather than just linking the region's (average) degree centrality to the number of its actors. To further investigate this issue, we take a closer look at how collaborative linkages (ties) are distributed between the *BioRegions*. Table 3 shows in a column-by-column manner the relative importance of each pairwise link, where row entries for each column add up to 1 (=100 percent). To give an example, Berlin-Brandenburg (No. 1) has 10 percent of all of its (intra- and interregional) collaborative research linkages with Bremen (No. 2), while the relative importance of collaborations with Berlin-Brandenburg from the perspective of Bremen is only 3 percent. Thus, the relative importance of interregional collaboration activity is not symmetric for the different *BioRegions*.

Another interesting result from Table 3 can be found if we look at entries on the diagonal (values in bold type), which display the percentage of collaborative linkages defined as loops. There are strong differences between the *BioRegions* as well. For instance, regional actors of the BioRegion Jena (No. 6) have –on average– about 13 percent of their linkages with partners within the region, while 33 percent of all linkages are between Jena and actors outside any established *BioRegion* ("99"-node), and roughly 17 percent are between Jena and the BioRegion Halle-Leipzig. This indicates that the role of interregional linkages may be more important compared to intraregional linkages, especially for small *BioRegions*. In comparison to Jena, the share of intraregional collaborations for the large BioTOP Initiative Berlin-Brandenburg is 25 percentage points higher (in total 38 percent). This large difference indicates that Berlin-Brandenburg has a sufficiently large internal absorptive capacity for R&D collaborations, while the BioRegion Jena heavily depends on external research partners. In Section 3, we will investigate whether information on the regional innovation system can be used to explain differences in absorptive capacity beyond core factors such as the number of biotech actors within the region.

Table 3. Relative importance of internal and external linkages for *BioRegions*

No.		1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.
	BioRegion	**10**	**99**	**14**	**2**	**13**	**9**	**8**	**15**	**16**	**7**	**1**	**5**	**4**	**3**	**12**	**6**	**17**
1.	**10**	**0.14**	0.06	0.04	0.05	0.04	0.04	0.06	0.03	0.01	0.04	0.03	0.01	-	0.02	0.08	0.03	-
2.	**99**	0.44	**0.31**	0.37	0.36	0.33	0.38	0.50	0.31	0.31	0.37	0.24	0.42	0.37	0.34	0.35	0.33	0.36
3.	**14**	0.04	0.05	**0.11**	0.05	0.07	-	0.01	-	0.04	0.11	0.06	0.03	-	0.02	-	-	-
4.	**2**	0.04	0.04	0.04	**0.23**	0.01	0.07	0.04	0.04	0.02	-	0.03	0.01	-	0.02	-	0.03	0.07
5.	**13**	0.07	0.08	0.12	0.02	**0.25**	0.07	0.01	0.07	0.05	0.04	0.06	0.07	0.09	0.03	-	-	-
6.	**9**	0.02	0.02	-	0.03	0.02	**0.13**	0.01	-	-	0.04	0.02	0.01	0.03	0.02	0.04	-	0.07
7.	**8**	0.07	0.08	0.01	0.05	0.01	0.02	**0.10**	0.07	0.04	0.04	0.03	0.05	-	0.03	0.08	0.03	-
8.	**15**	0.03	0.05	-	0.05	0.04	-	0.07	**0.26**	0.07	0.07	0.04	-	-	0.02	-	-	-
9.	**16**	0.01	0.05	0.05	0.02	0.03	-	0.04	0.07	**0.19**	0.04	0.03	-	0.09	0.11	-	0.13	0.14
10.	**7**	0.01	0.01	0.03	-	0.01	0.02	0.01	0.02	0.01	**0.11**	0.01	-	-	-	-	-	-
11.	**1**	0.09	0.12	0.21	0.10	0.13	0.16	0.08	0.12	0.11	0.15	**0.38**	0.13	0.06	0.15	0.12	0.10	-
12.	**5**	0.01	0.04	0.02	0.01	0.03	0.02	0.04	-	-	-	0.03	**0.16**	0.03	0.02	0.04	0.17	-
13.	**4**	-	0.02	-	-	0.02	0.02	-	-	0.03	-	0.01	0.01	**0.31**	0.02	-	-	-
14.	**3**	0.01	0.03	0.01	0.01	0.01	0.02	0.02	0.01	0.06	-	0.03	0.01	0.03	**0.18**	0.04	-	0.07
15.	**12**	0.02	0.01	-	-	-	0.02	0.02	-	-	-	0.01	0.01	-	0.02	**0.27**	-	-
16.	**6**	0.01	0.01	-	0.01	-	-	0.01	-	0.04	-	0.01	0.07	-	-	-	**0.13**	0.07
17.	**17**	-	0.01	-	0.01	-	0.02	-	-	0.02	-	-	-	-	0.02	-	0.03	**0.21**

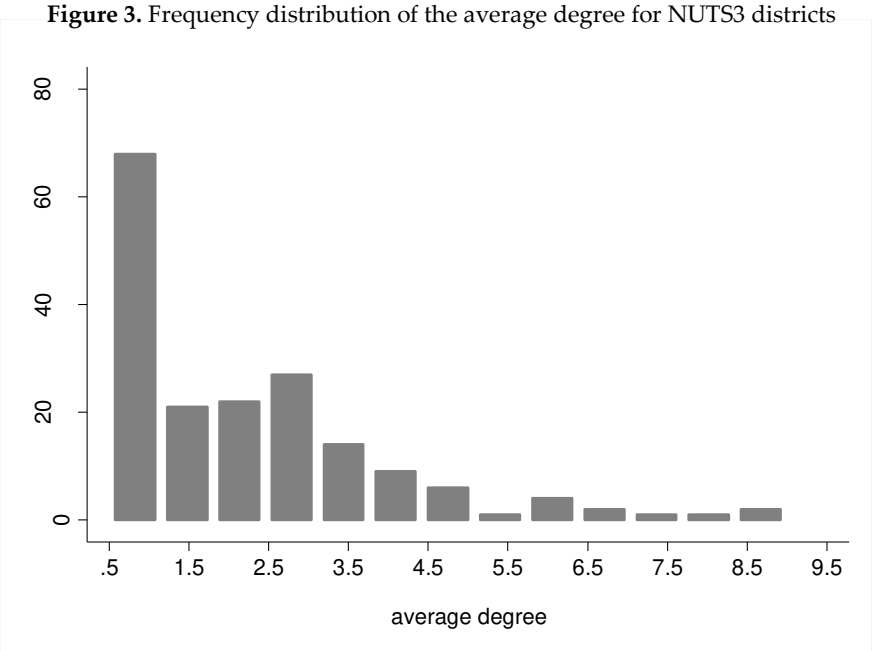
Notes: Calculated on the basis of data from BIOCOM (2005); listed *BioRegions* are numbered as in Figure 2.

With respect to the geographical distribution of collaboration activities across *BioRegions*, a striking fact is that actors in Jena tend to collaborate with external partners in close geographic proximity (Halle-Leipzig). Although Halle-Leipzig is a rather small regional cluster (28 actors), the relative share of collaborations with regional actors in Jena is much higher compared to the relatively large *BioRegions* such as Berlin-Brandenburg (133 actors, 10 percent). Thus, geographic proximity seems to matter for collaborative linkages among the *BioRegions*. The explicit role of geographic distance is also underlined by the fact that in nearly all cases, the highest weight is given to internal collaboration (values on the diagonal), even though there is a remarkable difference among *BioRegions*, ranging from a low value of 10 percent (Initiativkreis Biotechnologie München, No. 8) to a high value of 31 percent (BioRegion GreifswaldRostock, No. 4) or even 38 percent (BioTOP-Initiative Berlin-Brandenburg).

All in all, the qualitative inspection of the SNA-based indicators already shows that collaborative research behavior varies considerably among *BioRegions*. This heterogeneity can also be seen if we use the level of individual NUTS3 regions (rather than *BioRegions*) as the nodes of the intra-German collaboration network and plot the distribution of the average degree for the 178 NUTS3 regions with at least one biotech actor in the BIOCOM industry directory. The reader should note that the disaggregated NUTS3 level will also constitute the unit of analysis for a quantitative regression approach in the next section relating collaboration structures with regional attributes (membership in a *BioRegion* will then be measured by a set of binary dummies). As the histogram in Figure 3 shows, there is a wide range of average degree values at the regional level: In 68 out of the 178 regions (with at least one biotech actor) the average degree for the 178 NUTS3 regions is between

zero and one, while some districts even show values up to nine (that is, in these NUTS3 regions biotech actors have on average about 9 collaborative research linkages).

Figure 3. Frequency distribution of the average degree for NUTS3 districts

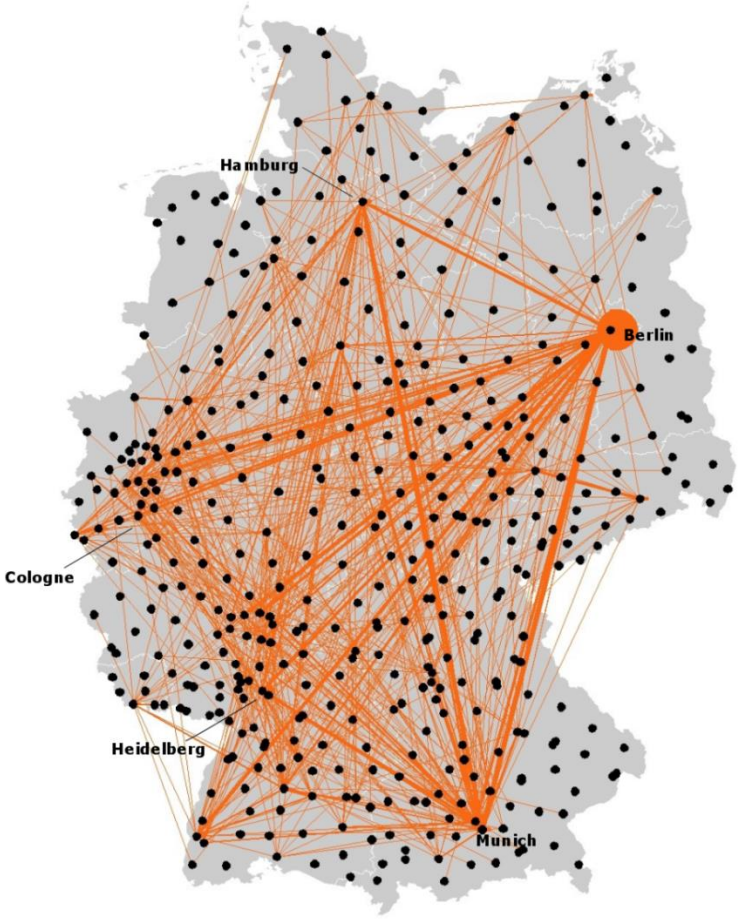


Note: Calculated on the basis of data from BIOCOM (2005) for 178 NUTS3 districts.

Finally, this regional diversity can also be gathered graphically if we draw a map of the biotech network at the NUTS3 level combining information on the linkages between the districts and their geographical location in Germany. The results in Figure 4 thereby show that large agglomerations, such as Berlin, Hamburg, and Munich, are at the core of the network and constitute important hubs through which collaboration activity runs even over large geographical distances. However, besides these dominant nodes also a wide range of smaller districts, such as Heidelberg, Tübingen, and Freiburg, show a strong performance in terms of cooperative behavior and network centrality. Since all of the latter actors are embedded in *BioRegion* cluster initiatives, this advocates the need to

elaborate an exploratory quantitative model of network formation accounting for the role played by the sectoral and regional innovation system.

Figure 4. Spatial distribution of collaborative linkages among German NUTS3 districts



Notes: Black dots mark centroids of the 439 German NUTS3 districts; orange dots and lines measure intra- and interregional collaboration activity; calculated on the basis of data from BIOCOM (2005).

3. Modeling interregional research collaborations the German biotech network

3.1 Econometric specification

When testing for the role played by regional and policy variables in determining link formation between nodes in the German biotech research network, we combine the descriptive SNA from above with an econometric modeling approach. As Knoke and Yang (2008) have pointed out, the field of SNA has steadily advanced with regard to statistical tools, such as exponential random

graph models (ERGM), for analysing tie formation as a function of node properties and the relationship of nodes towards each other. We will focus on econometric models for count data using the degree of research collaboration between each (i,j)-region tuple in the overall collaboration matrix of German NUTS3 regions with $i,j = (1,\dots,439)$ as outcome variable of interest. Since these collaborative linkages are undirected, our specified matrix has symmetric entries for two cells, (i,j) and (j,i).

In order to avoid an over-precision bias due to double counts, for estimation purposes we therefore only rely on observations in the lower triangular part of this collaboration matrix. Linkages (i,i) along the diagonal are covered as intraregional loops. This leaves us with a total number of $(440 \times 439)/2 = 96{,}580$ observations on collaboration activities at the level of NUTS3 districts. One advantage of our chosen dyadic (i,j)-specification compared to an aggregated analysis at the regional level (as for the degree centrality shown in eq.(1)) is that we have a higher number of observations, which increases estimation efficiency. Additionally, the pairwise estimation approach for each (i,j)-region tuple allows us to explicitly test for the influence of geographic distance between NUTS3 districts on the collaborative behavior of biotechnology actors which is typically assumed in the literature on proximity mechanisms to drive collaboration strength and network formation (see, e.g., Boschma, 2005; Ter Wal, 2014).

Given that we cover the entire geographical space of all German NUTS3 regions, one challenge for the estimation approach is that we have to deal with a large number of zero observations since only 178 of all 439 NUTS3 districts are found to host at least one biotech actor registered in the BIOCOM Year and Address book for 2005. Thus, besides applying standard count data specifications,

such as the Poisson or negative binomial regression models, we also test for the necessity to account for the inflation of zero entries in the number of collaboration activities by means of a zero-inflated Poisson (ZIP) or zero-inflated negative binomial (ZINB) specification. Zero-inflated models generally assume different data generating processes to be in order when predicting the probability for having no collaboration activity versus a positive collaboration activity on the one hand, and differences in the actual (non-zero) extent on the other hand.

The probability part of the model is estimated as a binary choice model (here: logit), which is then mapped into a standard Poisson or negative binomial specification. To guide model selection, different statistical tests will be used: Specifically, to judge whether a standard Poisson distribution with equal mean and variance is valid compared to a negative binomial model with under- or overdispersion in the expected mean value, we use a standard Likelihood Ratio (LR) test for the statistical significance of the overdispersion parameter in the empirical model. Likewise, a Vuong (1989) test will be applied to discriminate between the standard and zero-inflated specification (either for the Poisson or the negative binomial specification).

For the most general case of a zero-inflated negative binomial specification (ZINB), we model the number of collaborative linkages between i and j $\left(collab_{ij}\right)$ as

$$\log(\mu_{ij}) = \mathbf{X}_{ij}\beta' + u_{ij} \quad \text{and} \quad \text{logit}\left(\pi_{ij}\right) = \mathbf{Z}_{ij}\gamma' + \varepsilon_{ij}. \tag{2}$$

The ZINB assumes that there are two distinct data generation processes: The first part of eq.(2) describes the negative binomial part of the ZINB, which relates the conditional mean of the outcome variable $collab_{ij}$ defined as $E\left(collab_{ij}|\mathbf{X}_{ij}\right) = \mu_{ij}$ (for the case of a non-excessive zero) to a vector

of explanatory factors $\mathbf{X}_{ij}$**.** As shown in eq.(2), the logarithm (log) is used as the link function. In the binary choice part of the model, π_{ij} measures the probability that $collab_{ij}$ has excessive zero entries. The latter is related through a logit function to a vector of explanatory variables, $\mathbf{Z}_{ij}$. We thereby allow for the case that $\mathbf{X}_{ij}$ and $\mathbf{Z}_{ij}$ may contain the same set of variables. Finally, u_{ij} and ε_{ij} are the residuals in the negative binomial and binary choice part, respectively, where $\exp(u_{ij}) \sim Gamma(1/\tau, \tau)$ and $\varepsilon_{ij} \sim N(0, \sigma^2)$ with τ being the shape parameter quantifying the amount of overdispersion. Empirical estimation of the model according to eq.(2) is conducted by means of Maximum Likelihood (ML) technique (for a formulation of the (log) likelihood function of the ZINB, see, e.g., Mwalili et al., 2008).

3.2 Variable selection

The empirical literature on modeling research collaborations and network formation is fastly evolving (see, for instance, Hazir and Autant-Bernard, 2014; Wanzenböck et al., 2014, 2015; Broekel, 2015 for recent contributions). Determinants of network formation and the creation of new ties can thereby be broadly classified as either being related to capacity-based attributes of the underlying nodes (NUTS3 districts) or to different dimensions of the relationship between two nodes. Along these lines, we define three categories of variables for the estimation of eq.(2), which comprise

i.) "core" factors of the sectoral innovation system in biotechnology,

ii.) "capacity-based" indicators of the overall regional innovation system, and

iii.) "policy" variables related to signaling effects of the *BioRegio* contest and R&D funding (both through *BioRegio* and other channels of policy support in the biotech industry).

With regard to the core factors, in first place, we include the number of biotech firms in the region as a necessary condition for link formation. Moreover, we include the number of loops within NUTS3 regions as a separate determinant for the extent of interregional collaboration activity. We expect that the number of biotech patent applications is positively correlated with the region's ability to form collaborative linkages within the sector. As a third factor, we include the geographic distance between NUTS3 districts as an additional regressor, which can be motivated by a large empirical literature on the role of different proximity measures as determinants for network formation. In this literature, network relationships are supposed to be more common over short than long distances (Maggioni et al., 2007).

As the SNA has already shown, the intensity of intraregional collaboration within *BioRegions* is generally higher compared to interregional collaboration. Moreover, among interregional collaboration activities, regional actors tend to choose external partners nearby. There is ample empirical evidence on the role of proximity measures, particularly geographical distance, for network formation. Here, the literature predominantly finds that the quality and quantity of knowledge flows are subject to distance decay (Jaffe et al., 1993; Breschi and Lissoni, 2009) and that being located in clusters of spatial proximate actors has a positive impact on the actors' innovative performance (Baptista and Swann, 1998; Fornahl et al., 2011). At the EU regional level, Maggioni et al. (2007) as well as Scherngell and Barber (2009) provide evidence for the role of spatial proximity on interregional knowledge flows measured by co-patents and research projects funded by the EU Framework Programme(s). The latter authors find strong evidence for technological proximity among EU regions as a determinant for establishing research collaborations.

Capacity-based indicators stress the availability of (regional) resources as a significant contributing factor to absorb, exploit, and assemble different types of knowledge in the conduct of research and innovation activities (Cohen and Levinthal, 1990; Herrera and Nieto, 2008; Broekel and Brenner, 2011; Wanzenböck et al., 2014; Broekel, 2015). Since we treat NUTS3 regions as the nodes of the biotechnology network here, the large literature on regional innovation systems (RIS) is used to construct indicators on the basis of a region's human capital endowment, entrepreneurial activity, and international openness as well as localization and urbanization forces, which are likely to impact the regions' collaboration capacity (McCann, 2013). To test for the different channels through which the surrounding regional innovation system can impact the degree of collaboration activities between regions, we use a broad set of regional control variables, which are summarized in Table 4. Summary statistics for these variables are given in Table B.1 in the appendix.

Policy variables may have a distinct impact on network formation: Looking at the biotech industry in Germany, we find that public funding takes a prominent role, and the *BioRegio* contest in the mid-1990s can be seen as a "kick-off" event for massive policy support. As outlined above, the underlying intention of the *BioRegio* contest and its successors was to support regions with the best chances of success (Engel and Heneric 2005; Dohse 2007; Engel et al., 2013). Hence, if a *BioRegion* gains public support, the underlying collaboration activity of a region being member of an awarded *BioRegion* is likely to increase for two reasons. Firstly, there can be a direct effect as mentioned above: That is, the *BioRegion* has already had a noteworthy collaboration basis when starting to compete for financial support. And secondly, an awarded *BioRegion* might be seen as an interesting location for

finding new collaboration partners. Regarding the latter effect, being a winner (having high chances of succeeding in the biotech industry) might work as a signal for other actors in the industry.

However, the local network can also have a tendency for closure, and one may expect that while intraregional linkages increase due to *BioRegio* funding, interregional collaborative linkages are less likely to occur (see, for instance, Broekel et al., 2015). To capture these different effects, we therefore include variables that are directly related to R&D funding volumes in biotechnology as well as a set of binary dummies to reveal whether combinations of different categories of *BioRegio* participants show additional effects. The basic idea of including these dummy variables is to check for level differences between regions within and regions outside the *BioRegio* network, which may be related to preferential access to R&D funding and signaling/mobilizing effects of the *BioRegio* contest. We construct three types of dummies as

i.) *BioRegio (Winner x Winner):* The dummy variable takes the value of 1 for region pair (*ij)* if both NUTS3 regions are members of an awarded *BioRegio* cluster initiative and is 0 otherwise;

ii.) *BioRegio (Winner x Participant):* The dummy variable takes the value of 1 for region pair (*ij)* if one NUTS3 region is a member of an awarded *BioRegio* cluster initiative and the other NUTS3 region is a member of a non-winning *BioRegio* cluster initiative; the dummy is 0 otherwise;

iii.) *BioRegio (Participant x Participant):* The dummy variable takes the value of 1 for region pair (*ij)* if both NUTS3 regions are member of a non-winning *BioRegio* cluster initiative and is 0 otherwise.

Table 4. Variable definitions and source information

Variable	Description	Period	Source
Collaborative linkages	Number of total R&D collaborations between region *i* and *j* (including loops)	2005	BIOCOM AG
Actors	Number of biotech firms and institutions in region *i*	2005	BIOCOM AG
Loops	Number of total R&D collaborations of biotech actors within region *i*	2005	BIOCOM AG
Geographical distance	Driving time (in minutes) between the centroids of region *i* and *j*	2005	Federal Institute for Research on Building, Urban Affairs and Spatial Development (BBSR)
Biotech patent applications	Weighted number of patent applications in biotechnology in region *i* (OECD definition)	Sum of 1997-2002	European Patent Office (EPO)
Individual R&D Funding	Direct funding of biotechnology-related R&D individual projects by federal government in region *i* (in 1000 €)	Sum of 1997-2002	Projektförderungs-Informationssystem (PROFI)
Collaborative R&D funding	Direct funding of biotechnology-related R&D collaborative projects by federal government in region *i* (in 1000 €)	Sum of 1997-2002	Projektförderungs-Informationssystem (PROFI)
High-tech start-ups	Number of start-ups in high-tech industries relative to MINT employees in region *i* (1 = 100%)	Average of 1996-2003	ZEW Foundation Panel
Medium-tech start-ups	Number of start-ups in medium-tech industries relative to MINT employees in region *I* (1 = 100%)	Average of 1996-2003	ZEW Foundation Panel
KIS start-ups	Number of start-ups in knowledge-intensive services relative to MINT employees in region *i* (1 = 100%)	Average of 1996-2003	ZEW Foundation Panel
International openness	Share of foreign turnover in manufacturing sector relative to total turnover in the sector in region *i* (1 = 100%)	Average of 1997-2002	German Statistical Office
MINT employment	Share of employees trained in mathematics, informatics, natural sciences and technology in total employment of region *i* (in %)	Average of 1997-2002	Federal Employment Agency
Population density	Number of inhabitants per area in region *i* (in square kilometers)	Average of 1997-2002	German Statistical Office
Sectoral specialization manufacturing	Sum of squared deviations in employment shares for NACE3 sectors between region *i* and national average	1998	Alecke et al. (2006)
Sectoral specialization business-rel. services	Sum of squared deviations in employment shares for NACE3 sectors between region *i* and national average	1998	Alecke et al. (2006)
Sectoral specialization household-rel. services	Sum of squared deviations in employment shares for NACE3 sectors between region *i* and national average	1998	Alecke et al. (2006)
Ellison-Glaeser index manufacturing	Employment in sectors with high Ellison-Glaeser index (>0.005) relative to total employment in region *i*	1998	Alecke et al. (2006)
Ellison-Glaeser index business-rel. services	Employment in sectors with high Ellison-Glaeser index (>0.005) relative to total employment in region *i*	1998	Alecke et al. (2006)
Ellison-Glaeser index household-rel. services	Employment in sectors with high Ellison-Glaeser index (>0.005) relative to total employment in region *i*	1998	Alecke et al. (2006)

Notes: See Table B.1 in Appendix B for summary statistics of the variables.

Except for the set of dummies and the geographical distance between two NUTS3 regions, which vary for each $(\boldsymbol{i},\boldsymbol{j})$-region tuple, all regressors are measured at the regional level. In order to construct a set of doubly-indexed regressors as shown in eq.(2), we follow the literature on Gravity models (of trade) and calculate average values for log-transformed variables such as $\boldsymbol{x_{ij}} = \boldsymbol{(x_i + x_j)/2}$ (see, e.g., Rose, 2004). Loops for intraregional collaborations within NUTS3 regions are specified as a vector of constant values for each region tuple. Finally, with regard to sample organization, we impose a lag structure for the transmission channels running from the regressors to the model's outcome variable in order to reduce the problem of reversed causality: That is, while we use observations in 2005 for our dependent variable, the number of loops and the number of biotech firms as a normalizing factor (variables are taken from the BIOCOM Year and Address book), we impose a time lag of at least three years for all other regressors (see Table 4).

To give a practical example: When taking a closer look at the relationship between public R&D funding and the degree of collaboration activity between NUTS3 districts, we expect that public funding positively influences the collaboration activity of regions. However, a high degree of collaborative linkages is also likely to increase the probability of raising further public funding in the future. Similarly, regional economic and institutional conditions may influence the results of national innovation policy (Herrera and Nieto, 2008). The latter feedback mechanism results in a reversed causality problem between the two variables. In order to minimize this problem, we only use funding volumes allocated throughout the period 1997-2002 to measure the correlation with collaborative linkages in 2005. The period 1997-2002 was chosen since it reflects the *de facto* funding period in the *BioRegio* contest. Similarly, we chose lag structures for the remaining regressors as shown in Table 4.

Another limitation related to our estimation approach stems from the fact that we only observe research collaborations for a single sample year and thus have to estimate the model in a cross-sectional fashion. This surely limits the interpretation of our estimation results in terms of "cause→effects" statements compared to empirical identification approaches developed on the basis of panel data estimators. However, although we cannot fully control for cross-sectional heterogeneity by introducing region-fixed effects, we make use of spatial filters as a surrogate for these region-fixed effects (Patuelli et al., 2012). In fact, controlling for spatial autocorrelation by means of spatial filtering allows us to capture omitted variables and unobserved spatial spillover effects that can lead to inconsistent or inefficient estimation results particularly when working with regional data (Anselin, 1988).

We apply an eigenvector-based spatial filtering approach developed by Griffith (2003) to account for the potentially uneven –and spatially correlated– regional distribution of collaboration activities in the biotech industry (see Figure B.1 in the appendix for a graphical overview of selected variables at the NUTS3 level). A main advantage of the spatial filtering approach compared to alternative spatial regression techniques is that the former does not require the assumption of normality or other estimation restrictions and can be straightforwardly applied to count data regression approaches. As starting point, we extract orthogonal and uncorrelated numerical components (eigenvectors) from a projection matrix of an exogenously specified spatial weights matrix, $\mathbf{W}$. For the latter, we employ a "rook-type" binary contiguity weighting matrix, which takes values of 1 if two NUTS3 regions share a common geographical border and has zero entries otherwise.

In order to apply the spatial filtering approach to dyadic regression specifications, we transform the underlying **W** matrix into a network weighting matrix, **C**, which extends the two-dimensional space for $(N \times N)$-regional tuples with $(i, j, |i \neq j; i, j = 1, \dots, N)$ of **W** to a four-dimensional space with $(N^2 \times N^2)$ possible linkages for $i, j, r, s | i \neq j; r \neq s; i, j = 1, \dots, N; r, s = 1, \dots, N)$. We thereby allow for both origin- and destination-related linkages as outlined in Chun and Griffith (2010). The extracted eigenvectors from **C** are then included as additional regressors in eq.(2). To reduce the total number of included eigenvectors in the regression equation, we follow Grimpe and Patuelli (2011) and first select a subset of candidate eigenvectors according to the following threshold: $MI(e_i)/max_i[MI(e_i)] > 0.25$, where $MI(e_i)$ is Moran's I (*MI*) indicator for spatial autocorrelation computed based on a generic eigenvector, e_i. We use a stepwise regression approach to exclude statistically insignificant eigenvectors in each regression setup. Finally, we test for the joint significance of the remaining eigenvectors by means of a Wald test. Taken together, although this approach does not fully alleviate the problems associated with the estimation of a cross-section vis-à-vis a panel econometric specification, it ensures obtaining the most robust estimation results under the given data restrictions (see Kristin and Fischer, 2015 for a similar application to the estimation of dyadic trade data).

3.3 Empirical results

The estimation results for different count data specifications are shown in Table 5. In columns I and II we first estimate a negative binomial (NegBin) model, which includes the "core" factors (number of biotech firms and geographical distance in column I) and subsequently adds further regressors in column II. Column III additionally estimates the full models on the basis of a ze-

ro-inflated negative binomial (ZINB) model according to eq.(2). As the regression results show, in all specifications we obtain a positive and statistically significant regression parameter for the number of biotech actors indicating the existence of a link between the number of actors and the extent of collaboration activity. In line with the literature reviewed above, we also find a negative correlation between geographical distance and the number of collaborative linkages indicating that spatial proximity matters for tie formation in the biotechnology network composed of NUTS3 regions. We also find a negative correlation between the number of loops within a region and the interregional collaboration activity indicating that NUTS3 districts with a sufficient mass of intraregional collaboration activities tend to close their network with regard to external collaborations. This latter result relates to recent findings reported in Broekel et al. (2015) showing that actors embedded in strong local clusters tend to be less intensively embedded in national research collaboration networks (measured in terms of collaborative R&D funding).

With regard to the role of policy variables, the specification in column I of Table 5 only includes the set of binary dummies. The results show that particularly the combinations of winner regions in the contest as well as winning and non-winning participants have a statistically significant higher collaboration activity compared to other region pair types. In particular the positive effect found for regional pairs comprising a *BioRegio* winner and a non-winning participant region may hint at a double dividend of the *BioRegio* contest comprising both a monetary funding effect as well as a signaling effect. However, if we additionally include the volume of R&D funding received in columns II and III as additional regressors, we see that the positive signaling effect of the contest cancels out, indicating that the overall funding effect can be reduced to the pure monetary dimension of

funding. The absence of additional signaling effects for non-winning participants, once we control for the monetary effect of funding, may reflect the tendency of awarded *BioRegions* to close their networks and form dense cliques of strongly interconnected actors (see Ter Wal, 2014).

With regard to the different types of R&D funding, we particularly get evidence for a positive correlation between the volume of collaborative R&D grants and the number of collaborative linkages, while the volume of individual R&D grants turns out to be statistically insignificant in the ZINB specification in column III. Our obtained results support the findings by Fornahl et al. (2011), namely that individual R&D subsidies do not enhance the performance of biotech firms in terms of patent activity, while collaborative research subsidies, in fact, do. Since we are using logarithmic transformations for the set of regressors, the obtained regression parameters can be interpreted in a straightforward manner as elasticities. Thus, a 1 percent increase in the volume of collaborative R&D funding leads to a 0.15-0.23 percent change in the number of pairwise collaboration linkages. Surprisingly, we do not get statistical evidence for a positive link between the number of biotech patent applications and the collaboration activity of region pairs at the NUTS3 level in the ZINB, while the variable is positive and statistically significant in the NegBin specification in column II.

Table 5. Estimation results for the determinants of collaborative links between NUTS3 regions

Dep. var.:	$collab_{ij}$	**NegBin (I)**	**NegBin (II)**	**ZINB (III)**
Core	Actors	0.085	0.041	0.042
		(0.007)***	(0.007)***	(0.007)***
	Loops	-0.009	-0.007	-0.007
		(0.003)***	(0.003)**	(0.002)***
	Geographic distance	-0.839	-0.751	-0.587
		(0.049)***	(0.047)***	(0.059)***
	Biotech patent applications		0.286	-0.058
			(0.045)***	(0.082)
Policy	Individual R&D funding		0.093	0.024
			(0.015)***	(0.027)
	Collaborative R&D funding		0.223	0.160
			(0.018)***	(0.040)***
	BioRegio dummy (winner x winner)	0.670	-0.005	0.227
		(0.289)**	(0.244)	(0.216)
	BioRegio dummy (winner x participant)	0.475	-0.134	0.068
		(0.194)**	(0.174)	(0.163)
	BioRegio dummy (participant x participant)	0.426	-0.275	-0.168
		(0.245)*	(0.230)	(0.224)
RIS	High-tech start-ups	0.720	0.528	0.544
		(0.166)***	(0.191)***	(0.198)***
	Medium-tech start-ups	-0.369	-0.384	0.028
		(0.205)*	(0.224)*	(0.246)
	KIS start-ups	2.405	0.369	1.138
		(0.309)***	(0.288)	(0.336)***
	International openness	-0.023	-0.102	-0.164
		(0.055)	(0.052)**	(0.054)***
	MINT employment	1.435	0.535	0.307
		(0.198)***	(0.201)***	(0.207)
	Population density	0.551	0.412	0.310
		(0.086)***	(0.084)***	(0.088)***
	Specialization manufacturing	-0.140	-0.035	-0.000
		(0.091)	(0.091)	(0.091)
	Specialization business-rel. services	-1.197	-0.244	-0.400
		(0.140)***	(0.133)*	(0.143)***
	Specialization household-rel. services	-0.322	-0.162	-0.342
		(0.107)***	(0.096)*	(0.102)***
	Ellison-Glaeser index manufacturing	-0.928	-0.548	-0.641
		(0.103)***	(0.097)***	(0.105)***
	Ellison-Glaeser index business-rel. services	-0.468	0.029	-0.313
		(0.283)*	(0.284)	(0.299)
	Ellison-Glaeser index household-rel. services	0.438	0.168	0.410
		(0.223)**	(0.223)	(0.235)*
No. of observations		96,579	96,579	96,579
Wald test (χ^2) for spatial filters		372.69***	302.22***	271.30***
LR-test (Poisson vs. NegBin)		177.17***	92.54***	4.26**,$
Vuong (NegBin vs. ZINB)				2.38***

Notes: * p<0.1; ** p<0.05; *** p<0.01; $ = LR-test for zero-inflated Poisson vs. ZINB. Standard errors are given in brackets. A constant term has been included in the regression but is not reported here. See main text for details.

Regarding the impact of factors from the regional innovation system, the ZINB results show that cooperative behavior is positively correlated with the share of business start-ups in high-tech sectors and knowledge-intensive services (KIS) according to the OECD (2010) classification. While we also observe a positive regression coefficient for general agglomeration factors (measured in terms of population density), localization forces (proxied by industry specialization and sectoral concentration) generally show a negative correlation with the number of regional linkages. This latter finding can be brought in line with Cantner and Graf (2004), who argue that for high-tech regions the collaboration activity is expected to be the highest for some intermediate degree of specialization. Similarly, we find a negative correlation between international openness and the number of research collaborations.

With regard to post-estimation tests for model selection, the ZINB specification can also be seen as the preferred empirical choice on the basis of the reported Vuong test. Additionally, as shown in Table 5, we get empirical evidence for the role played by overdispersion favoring the negative binomial distribution against the Poisson distribution (LR-test). As the post-estimation tests further show, the included spatial filter turns out to be statistically significant in all regression specifications indicating that it is important to account for spatial network effects.

Finally, the results for the underlying binary choice part of the ZINB reported in Table 6 indicate that both individual and collaborative R&D funding as well as the number of biotech patent applications are essential for having collaborative linkages at all. Thus, having a critical mass in research activity –as proxied by the number of patent applications– and having access to public R&D grants may be seen as important prerequisites for engaging in research collaborations. The

negative role of distance also turns out to be statistically significant in the logit part of the ZINB model. This means that increasing the distance between two regions lowers the probability that any interregional collaboration will occur. The reader should note that we have only included core sectoral and policy variables in the first-step logit model specification in Table 6. However, these results remain stable even if we include the full set of variables as in Table 5.

Table 6. Estimation results for the binary choice part (logit) of the ZINB specification

Dep. var.:	π_{ij}	**ZINB**
	Geographic distance	0.313
		(0.105)***
	Biotech patent applications	-0.426
		(0.094)***
	Individual R&D funding	-0.117
		(0.039)***
	Collaborative R&D funding	-0.094
		(0.054)*

Notes: * p<0.1; ** p<0.05; *** p<0.01. Standard errors are given in brackets. A constant term has been included in the regression but is not reported here. Results from alternative logit model specifications can be obtained from the authors upon request.

3.4 Robustness tests

This section serves as a critical appraisal of the use of the BIOCOM industry directory data as a novel source of information to model the German research collaboration network in biotechnology. To do so, we first compare its structure with the patent citation network at the regional level. Patent citations are proposed in the literature as a widely accepted measure for knowledge flows within a sectoral innovation system. As Breschi and Lissoni (2004) point out, the reason for using patent citations as a meaningful indicator for knowledge flows resides in the view of innovation as a social process implying that inventors often need to exchange tacit knowledge with other inventors beyond using bibliographic sources and personal experiments.

Following the classification of biotech-related IPC classes as outlined in Lecocq (2010), we count sector-specific patent citations between German NUTS3 regions on the basis of the OECD RegPAT database (Maraut et al., 2008). Geographic information on the citing and cited patent is retrieved at the applicant level. Moreover, directed patent citations from region *i* to region *j* and from region *j* to region *i* have been summed to get an undirected measure of two-way patent citations for region pair *(i,j)*, which comes closest to the undirected collaboration measure from the BIOCOM industry directory. With regard to the sample period, we select all biotech-related patent applications at the European Patent Offices (EPO) between 1997 and 2005, which can be expected to reflect the collaboration network in German biotechnology until 2005 in the context of the *BioRegio* contest starting in 1997. Besides the comparison across indicators, we also gather collaboration data from the BIOCOM Year and Address book 2009 in order to check for variations and discontinuities in the industry directory data over time. We use the same data aggregation procedure as outlined above.

While Table 7 reports summary measures for each of the three networks (Patent Citations 1997-2005, BIOCOM 2005, and BIOCOM 2009) together with the results of a bivariate correlation analysis, Figure 5 provides a graphical presentation of the two alternative networks in similar veins as Figure 4 for the BIOCOM research collaboration network in 2005. When we compare the summary measures, the results show that the collaboration networks based on the BIOCOM data for 2005 and 2009 have a significantly larger average degree compared to the EPO patent citations network. This difference is also highlighted by the lower number of interregional linkages for patent citations in Figure 5. However, while this level difference can mainly be attributed to the specific characteristics of patent citations as one specific channel of research connectivity, the link structure for the patent citation network and BIOCOM collaboration networks (both in 2005 and 2009) is very similar highlighting the role played by key linkages connecting the large regional hubs of the biotech sectors such as Berlin, Munich, Heidelberg, and the Rhineland. This rectified structure is also reflected by the pairwise correlation coefficients for the patent citation network and the BIOCOM collaboration networks, which show positive and statistically significant coefficients of 0.30 (2005)

and 0.31 (2009), respectively. Finally, with regard to the average clustering coefficient and the average path length, the BIOCOM collaboration and EPO patent citation networks show to have similar values as well.

Table 7. Comparison of the alternative networks of interregional research collaborations in German biotech

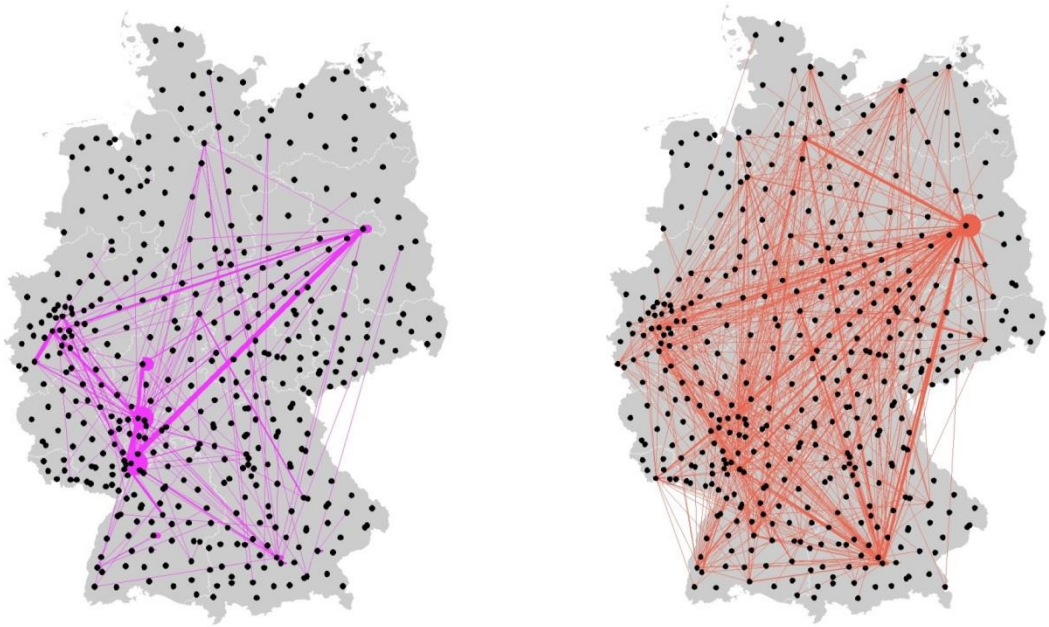
Indicator	BIOCOM collaborations 2005	BIOCOM collaborations 2009	EPO Patent citations 1997–2005
Average degree	6.351	6.556	1.021
Average clustering coefficient	0.373	0.302	0.267
Average path length	2.774	2.822	3.241
Pearson's correlation coefficient			
BIOCOM collaborations 2005	1		
BIOCOM collaborations 2009	0.88***	1	
EPO Patent citations 1997–2005	0.31***	0.30***	1

Notes: * $p<0.1$; ** $p<0.05$; *** $p<0.01$.

When we take a closer look at the change in the BIOCOM network between 2005 and 2009, the highly positive and statistically significant correlation coefficient reported in Table 7 indicates that interregional collaboration activity is quite persistent over time. The further network indicators in Table 7 moreover show that the average degree for distinct regions in the network has grown between 2005 and 2009 indicating that regions have still increased their outreach in terms of research collaborations with other regions throughout the mid-2000s. At the same time, the average clustering coefficient as a measure to which degree nodes cluster together has decreased between 2005 and 2009, while the average path length has increased slightly. These reversed trends may reflect changes in the organization of the sectoral innovation system, for instance, driven by shifts in the funding regime with a weaker focus on network and cluster policies. The latter trend can also be identified when we re-estimate the model from eq.(2) using the BIOCOM collaboration network in 2009 as the outcome variable and add the interregional collaboration structure in 2005 as an additional regressor (a detailed regression output is given in Table B.2 in the appendix).

Figure 5. Alternative measures for the interregional research collaboration network of German NUTS3 districts

(a) EPO Patent Citations 1997–2005 *(b) BIOCOM Industry Directory 2009*



Notes: Black dots mark centroids of the 439 German NUTS3 districts; colored dots and lines measure intra- and interregional collaboration activity; for details on calculation see main text.

Although most regression coefficients remain stable, two noteworthy findings can be gathered: Firstly, different from the network structure in 2005, the degree of interregional research collaboration (in 2009) is not significantly correlated with previous levels of collaborative R&D funding, while the spatial distribution of individual R&D funding volumes appears to matter. This result may hence reflect the above-mentioned shift in the overall funding regime in biotechnology. Secondly, the estimation results in Table B.2 show that the coefficient for the collaboration activity in 2005 enters the model for 2009 in a positive and statistically significant way hinting at temporal autocorrelation in the interregional collaboration activity over time. Although this latter result has to be interpreted carefully given that it may simply reflect the tendency of actors to seek (and/or report) stable long-run collaborative linkages, at the regional level the result may also be interpreted in favor of a temporal network evolution in the form of preferential attachment – that is, nodes with an above average collaboration degree attract new links at a higher rate over time compared to nodes

with an initially lower collaboration degree (Barabasi and Albert, 1999). As Fritsch and Kudic (2016) point out, this development may be particularly relevant at the system level of networks.

4. Discussion

The findings from our empirical investigation have theoretical and practical implications for the emerging field of innovation studies, which aim at modeling research collaboration activities from a network perspective. Additionally, several policy implications emerge, which will be discussed in the following subsections.

4.1 Theory and Practice

Theoretical research on network formation has grown considerably in recent years. However, when it comes to the empirical modeling of knowledge networks and its main drivers related to node characteristics and the relationship between nodes, scholars often face binding data limitations. In this paper, we have proposed the use of commercial industry directory data as a novel source of information to map and model the interregional research collaboration activity for German biotechnology. Further, we have placed a particular focus on the interregional dimension of collaboration given that the biotech sector is typically characterized by the emergence of urban centers and spatial clusters, which act as knowledge hubs in the sectoral innovation system.

Using indicators commonly employed in the field of social network analysis together with an econometric modeling approach for count data, we have then tested theoretically motivated model predictions related to the role of proximity dimensions for network formation and were able to study the temporal evolution of networks linked to an assessment of the role played by preferential attachment mechanisms. Our results support earlier findings hinting at the role of geographic distance as an impediment to collaboration. They also hint at the role played by modern location factors for the formation of network ties. With regard to robustness tests, we find that the BIOCOM industry directory data on research collaboration are significantly correlated with the biotech patent citation

network of German regions and can thus be seen as a complementary indicator to map and model knowledge flows in German biotechnology.

4.2. Policy

Our empirical results have also shown that public policy inputs are significantly related to the strength of interregional research collaborations – mainly through monetary incentives in terms of collaborative R&D support schemes during the *BioRegio* contest. Our focus on this policy instrument can be justified by its prototype role for the larger family of national, sector-specific, and multi-sectoral cluster policy programs in Germany. Examples of these policies include the *BioProfile* and *BioChance* funding schemes within the biotech sectors, as well as *InnoRegio* and the *Spitzencluster-Wettbewerb* for multi-sectoral funding schemes. Moreover, several cluster policy instruments at the regional level have adopted the original idea of the *BioRegio* contest as well (for instance *BioRegio Bayern 2020* in Bavaria and *RegioWIN* in Baden-Württemberg). Particularly the *Spitzencluster-Wettbewerb* can be seen as an important instrument of the high-tech strategy of the German government (Cantner et al., 2015). Similar to the *BioRegio* contest, the *Spitzencluster-Wettbewerb* (2008-2014) has been based on a contest-of-cooperation approach and has been pursued in three rounds of competition with a total of 15 winner cluster initiatives from various industries, such as biotechnology, aviation, e-mobility, micro- and nano-electronics, organic electronics, etc.

Broadening the scope from an industry-specific cluster policy such as *BioRegio* to multi-sectoral policy programs has led to substantial heterogeneity among cluster initiatives in terms of formal structure, organization, and research activities (Cantner et al., 2015). Also the size of their geographical areas and distances within cluster boundaries differ substantially due to a fuzzier interpretation of the cluster concept. Accordingly, empirical results for network formation in the awarded cluster initiatives of the *Spitzencluster-Wettbewerb* were twofold*:* On the one hand, in line with our empirical results, Cantner et al. (2015) find that network formation is mainly driven by monetary incentives stemming from the cluster policy without any additional (signaling) impact. On the other hand, the authors also find a considerable heterogeneity in the observed effects across

industries. This indicates that policymakers should be cautious about adopting policy schemes that work properly within a specific sectoral context to other domains in a 1-1 fashion (see also Dolfsma and Seo, 2013).

5. Conclusions

This paper has conducted an empirical investigation of the structure and determinants of the research collaboration network in German biotechnology. One novelty of our approach rests on the use of –so far– largely unexplored commercial industry directory data as a source of information for mapping and modeling research collaborations in German biotechnology. Another novelty stems from our distinct focus on regional entities (NUTS3 regions) as nodes of the German research collaboration network. This focus can be motivated on the basis of the role that distinct urban centers and regional clusters play in the German biotech sector. From a methodological perspective, we have combined social network analysis and econometric count data modeling to study the strength of interregional research collaboration activity as a function of node properties (e.g., the underlying regional innovation system) and the relationship between nodes (e.g., geographic distance between regions). By doing so, our explorative empirical analysis was able to shed new light on the determinants of network formation within a sectoral innovation system. This, in fact, can be seen as highly relevant for policymakers interested in fostering the collaboration activity of research actors in the short to midterm as a means to cater the long-run policy goal of increasing the economy's innovative performance (Zeng et al., 2010; Kang and Park, 2012).

Our empirical analysis has shown that the structure of the research collaboration network in German biotechnology is far from being random and features specific factors that can be related to underlying regional research and economic endowments as well as the relationship between regions. As such, we find that modern locational factors are positively correlated with the extent of interregional research collaborations, while geographic distance is found to be an impediment to collaboration. Moreover, network and cluster policies, in particularly the *BioRegio* contest analysed

in the context of German biotech funding, appear to be a promising instrument of public policy as we find that the volume of collaborative R&D funding is positively correlated with the interregional collaboration activity among German NUTS3 regions. One should note, though, that the empirical approach chosen in this paper should merely be seen as an explorative modeling exercise rather than a strict "causal" impact analysis with regard to the economic effects of cluster policies (see, e.g., Falck et al., 2010, Martin et al., 2011, or Engel et al., 2013 for rigorous causal-impact studies). Future research should thus particularly focus on these latter tools when assessing the different long-run goals of policy seen from a network perspective. Additionally, future studies may further investigate the dynamic nature of network formation within sectoral innovation systems. The use of industry directory data as a complementary data source next to widely used patent indicators can be seen as a fruitful approach in this endeavor.

References

Alecke, B.; Alsleben, C.; Scharr, F.; Untiedt, G. (2006): Are there really high-tech clusters? The geographic concentration of German manufacturing industries and its determinants, in: Annals of Regional Science, 40(1): 19-42.

Anselin, L. (1988): Spatial Econometrics: Methods and Models, Kluwer: Dordrecht.

Balconi, M.; Breschi, S.; Lissoni, F. (2004): Networks of inventors and the role of academia: an exploration of Italian patent data, in: Research Policy, 33(1): 127-145.

Baptista, R.; Swann, G. (1998): Do firms in clusters innovate more?, in: Research Policy, 27(5): 525-540.

Barabási, A.; Albert, R. (1999): Emergence of scaling in random networks, in: Science, 386(15): 509-512.

Bathelt, H.; Malmberg, A.; & Maskell, P. (2004): Clusters and knowledge: Local buzz, global pipelines and the process of knowledge creation, in: Progress in Human Geography, 28: 31–56.

BIOCOM AG (2005): BioTechnologie – Das Jahr- und Adressbuch 2005, Mietzsch, A. (Ed.), 19. Jahrgang, Berlin.

BIOCOM AG (2009): BioTechnologie – Das Jahr- und Adressbuch 2005, Mietzsch, A. (Ed.), 23. Jahrgang, Berlin.

BMBF (2010): Ideas. Innovation. Prosperity. High-Tech Strategy 2020 for Germany, Bundesministerium für Bildung und Forschung (BMBF): Bonn und Berlin.

Boschma, R. (2005): Proximity and innovation: A critical assessment, in: Regional Studies, 39(1): 61-74.

Breschi, S.; Lissoni, F. (2009): Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows, in: Journal of Economic Geography, 9: 439-468.

Broekel, T.; Brenner, T. (2011): Regional factors and innovativeness: an empirical analysis of four German industries, in: Annals of Regional Science, 47: 169-194.

Broekel, T. (2015): Do cooperative R&D subsidies stimulate regional innovation efficiency? Evidence from Germany, in: Regional Studies, 49(7): 1087-1111.

Broekel, T.; Fornahl, D.; Morrison, A. (2015): Another cluster premium: Innovation subsidies and R&D collaboration networks, in: Research Policy, 44(8): 1431–1444.

Cantner, U.; Graf, H. (2004): Cooperation and Specialization in German Technology Regions, in: Journal of Evolutionary Economics, 14(5): 543-562.

Cantner, U.; Graf, H.; Töpfer, S. (2015): Structural dynamics of innovation networks in German Leading-Edge Clusters, Jena Economic Research Papers No. 2015-26.

Chun, Y.; Griffith, D. (2010): Modelling network autocorrelation in Space-Time Migration Flow Data: An Eigenvector Spatial Filtering Approach, in: Annals of the Association of American Geographers, 101(3): 523-536.

Cohen, W.; Levinthal, D. (1990): Absorptive capacity: a new perspective on learning and innovation, in: Administrative Science Quarterly, 35: 128-152.

Cooke, P.; De Laurentis, C.; Tödtling, F.; Trippl, M. (2007): Regional Knowledge Economies – Markets, Clusters and Innovation, Edward Elgar Publishing: Cheltenham.

Dohse, D. (2007): Cluster-Based Technology Policy – The German Experience, in: Industry and Innovation, 14(1): 69-94.

Dolfsma, W.; Seo, D. (2013): Government policy and technological innovation – a suggested typology, in: Technovation, 33: 173-179.

Engel, D.; Heneric, O. (2005): Biotechnologiegründungen im Ruhrgebiet - Eine vergleichende Analyse, RWI: Materialien, Heft 21, RWI: Essen.

Engel, D.; Mitze, T.; Patuelli, R.; Reinkowski, J. (2013): Does cluster policy trigger R&D activity? Evidence from German biotech contests, in: European Planning Studies, 21(11): 1735-1759.

Ernst & Young (2005): Kräfte der Evolution. Deutscher Biotechnologie-Report 2005, Ernst & Young AG Wirtschaftsprüfungsgesellschaft: Mannheim.

Falck, O.; Heblich, S.; Kipar, S. (2010): Industrial innovation: Direct evidence from a cluster-oriented policy, in: Regional Science and Urban Economics, 40(6): 574-582.

Fornahl, D.; Broekel, T.; Boschma, R. (2011): What drives patent performance in German biotech firms? The impact of R&D subsidies, knowledge networks and their location, in: Papers in Regional Science, 90(2): 395-418.

Freeman, L. C. (1978/79): Centrality in Social Networks – Conceptual Clarification, in: Social networks, 1(3): 215-239.

Fritsch, M.; Kudic, M. (2016): Preferential Attachment and Pattern Formation in R&D Networks – Plausible explanation or just a widespread myth?, Jena Economic Research Papers No. 2016-005.

Gertler, M.; Levitte, Y. (2005): Local Nodes in Global Networks: The Geography of Knowledge Flows in Biotechnology Innovation, in: Industry and Innovation, 12(4): 487-507.

Griffith, D. (2003): Spatial Autocorrelation and Spatial Filtering. Berlin.

Grimpe, C.; Patuelli, R. (2011): Regional Knowledge Production in Nanomaterials: A Spatial Filtering Approach, in: Annals of Regional Science, 46(3): 519-541.

Hazir, C.; Autant-Bernard, C. (2014): Determinants of cross-regional R&D collaboration: some empirical evidence from Europe in biotechnology, in: Annals of Regional Science, 53(2): 369-393.

Jaffe, B.; Trajtenberg, M.; Henderson, R. (1993): Geographic localization of knowledge spillovers as evidenced by patent citations, in: Quarterly Journal of Economics, 108(3): 577-598.

Kang, K.N.; Park, H. (2012): Influence of government R&D support and inter-firm collaborations on innovation in Korean biotechnology SMEs, in: Technovation, 32: 68-78.

Kesteloot, K.; Veugelers, R. (1995): Stable R&D Cooperations with Spillovers, in: Journal of Economics and Management Strategy, 4(4): 651-672.

Knoke, D.; Yang, S. (2008): Social Network Analysis, 2. edition, Sage: Thousand Oaks.

Lecocq, C. (2010): Technological Performance of Regions (and Firms). The Case of Biotechnology, PhD thesis, Faculty of Business and Economics, Katholieke Universiteit Leuven.

Ma, Z.; Lee, Y. (2008): Patent application and technological collaboration in inventive activities: 1980–2005, in: Technovation, 28(6): 379-390.

Maggioni, M.; Nosvelli, N.; Uberti, T. (2007): Space versus networks in the geography of innovation: European analysis, in: Papers in Regional Science, 86(3): 471-493.

Martin, P.; Mayer, T.; Mayneris, F. (2011): Public Support to Clusters: A Firm Level Study of French 'Local Productive Systems, in: Regional Science and Urban Economics, 41(2): 108–123.

Maskell, P.; Bathelt, H.; Malmberg, A. (2006): Building global knowledge pipelines: The role of temporary clusters, in: European Planning Studies, 14(8): 997-1013.

McCann, P. (2013): Modern urban and regional economics. 2nd edition, Oxford University Press.

Mwalili, S.; Lesaffre, E.; Declerck, D. (2008): The zero-inflated negative binomial regression model with correction for misclassification: an example in caries research, in: Statistical Methods in Medical Research, 17: 123-139.

OECD (2010): Science, Technology and Industry Scoreboard 2009, OECD: Paris.

Ó hUallancháin, B.; Lee, D.S. (2014): Urban centers and networks of co-invention in American biotechnology, in: Annals of Regional Science, 52: 799–823.

Paci, R.; Marrocu, E.; Usai, S. (2014): The complementary effects of proximity dimensions on knowledge spillovers, in: Spatial Economic Analysis, 9(1): 9-30.

Patuelli, R.; Schanne, N.; Griffith, D.; Nijkamp, P. (2012): Persistence of Regional Unemployment: Application of a Spatial Filtering Approach to Local Labour Markets in Germany, in: Journal of Regional Science, 52(2): 300–323.

Quintana-Garcia, C.; Benavides-Velasco, C. (2004): Cooperation, competition and innovative capability: a panel data of European dedicated biotechnological firms, in: Technovation, 24: 927-938.

Rose, A. (2004): Do we really know that the WTO increases trade?, in: American Economic Review, 94(1): 98-114.

RWI, SG, Joanneum Research und Friedrich-Schiller Universität Jena (2014): Begleitende Evaluierung des Förderinstruments „Spitzencluster-Wettbewerb" des BMBF: Abschlussbericht. RWI

Projektberichte, available at: http://www.rwi-essen.de/publikationen/rwi-projektberichte/ (last accessed: 18.01.2017).

Scherngell, T.; Barber, M. (2009): Spatial interaction modelling of cross-region R&D collaborations: empirical evidence from the 5th EU framework programme, in: Papers in Regional Science, 88(3): 531-546.

Schweitzer, F.; Fagiolo, G.; Sornette, D.; Vega-Redondo, F.; Vespignani, A.; White, D. (2009): Economic Networks: The New Challenges, in: Science, 325: 422-425.

Statistisches Bundesamt (2005): Unternehmen der Biotechnologie in Deutschland. Ergebnisse der Wiederholungsbefragung 2004, Statistisches Bundesamt: Wiesbaden.

Ter Wal, A. (2014): The dynamics of the inventor network in German biotechnology: geographic proximity versus triadic closure, in: Journal of Economic Geography, 14(3): 589-620.

Uyarra, E.; Ramlogan, R. (2012): The Effects of Cluster Policy on Innovation. Compendium of Evidence on the Effectiveness of Innovation Policy Intervention. Working paper, Manchester Institute of Innovation Research, Manchester Business School, University of Manchester.

Veugelers, R. (1998): Collaboration in R&D: an assessment of theoretical and empirical findings, in: De Economist, 146(3): 419-443.

Vuong, Q. (1989): Likelihood ratio tests for model selection and non-nested hypotheses, in: Econometrica, 57: 307-333.

Wanzenböck, I.; Scherngell, T.; Brenner, T. (2014): Embeddedness of region in European knowledge networks: a comparative analysis of inter-regional R&D collaborations, co-patents and co-publications, in: Annals of Regional Science, 53(2): 337-368.

Wanzenböck, I.; Scherngell, T.; Lata, R. (2015): Embeddedness of European regions in the European Union-funded research and development (R&D) network: A spatial econometric perspective, in: Regional Studies, 49(10): 1685-1705.

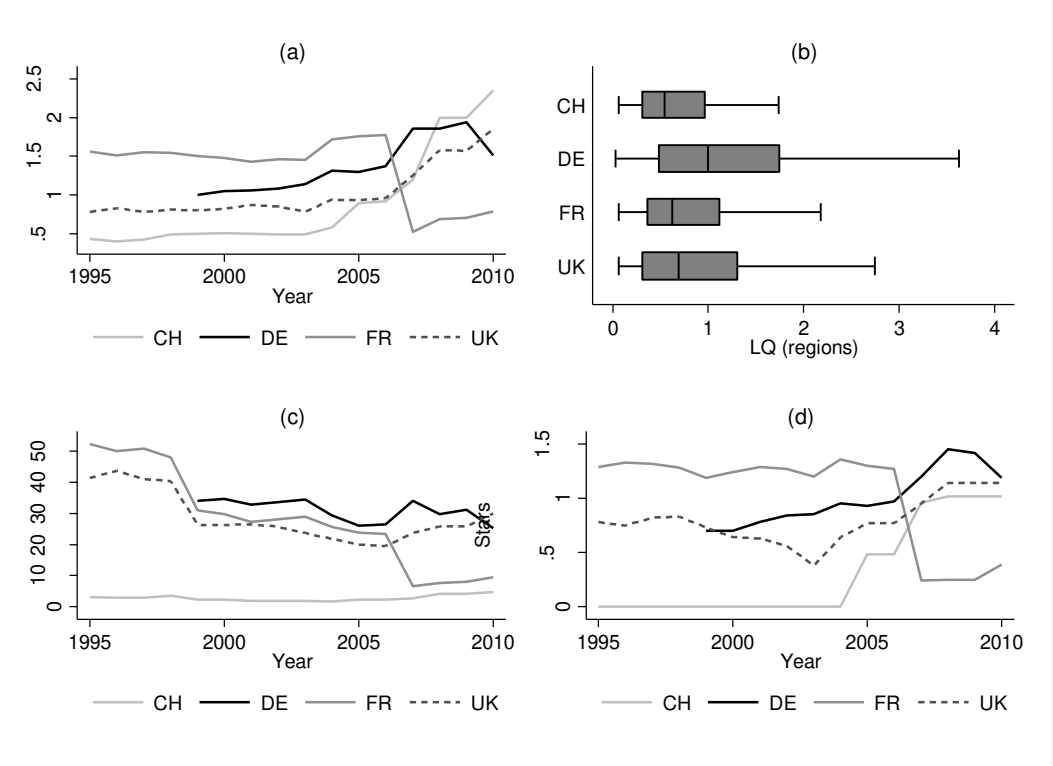
Zeng, S.X; Xie, X.M.; Tam, C.M. (2010): Relationship between cooperation networks and innovation performance of SMEs, in: Technovation, 30: 181-194.

Appendix A: Evolution of German biotechnology in an international context

While the German economy was characterized by having a merely inhospitable climate for biotechnology two decades ago (Dickman, 1996), today the country has developed to one of Europe's leading biotechnology countries. As Müller (2002) outlines, the number of dedicated biotechnology firms has grown from 17 to 333 in the period between the early 1990s and 2001 – thereby outpacing the number of biotech firms in the UK (271) and France (240). More recent data from the European cluster observatory (2016) show that the German biotechnology industry has manifested its leading role as national biotechnology player within Europe since then. Figure A.1 plots different indicators characterizing the national innovation systems in biotechnology with regard to industry concentration (location quotient, see Isserman, 1977), overall industry size as well as a composite "stars" indicator for the period 1995-2010. The European cluster observatory thereby defines the biotechnology industry according to the 4-digit level NACE Rev. 2 classification as 72.11 "Research and experimental development on biotechnology", which is in line with similar studies as in Laskawi (2015).

The location quotient (LQ) in panel (a) of Figure A.1 compares the extent to which the national economies of Germany, the UK, France, and Switzerland have an above average concentration of biotechnology employment relative to the EU-27 average (indicated by $LQ \geq 1$). As the figure shows, the industry concentration in Germany and the UK has been constantly above the EU-27 average for the sample period 1995-2010. In comparison, biotechnology-related employment concentration in Switzerland has rapidly grown in the second half of the last decade, while the LQ value for France has dropped considerably. Reasons for the striking decline in French biotechnology can be attributed to an increasing relocations of firms (particularly to Switzerland), a funding crisis with a drastic decline in equity investments and no active financial market as back-up (PharmaLetter, 2005a&b, BioSpace, 2009) combined with the worsening of the overall macroeconomic conditions in the course of the global financial and economic crisis (France Biotech, 2008).

Figure A.1. Evolution of biotechnology industry in selected European countries (1995–2010)



Notes: CH = Switzerland, DE = Germany, FR = France, UK = United Kingdom. The four panels display the following information: (**a**) LQ (country); (**b**) LQ (regions); (**c**) size and (**d**) observatory star rating. For details on the definitions of these indicators see main text; calculated on the basis of data from the European cluster observatory (2016).

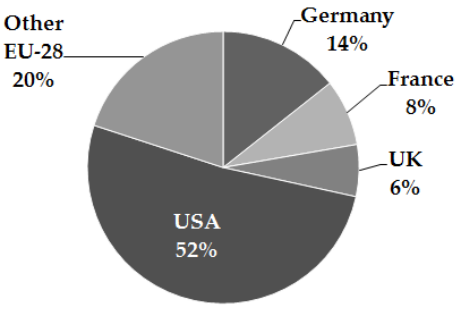
In addition to the computation of national LQ values for the biotechnology sector in the four economies, panel (b) in Figure A.1 shows a box plot graph of the intra-national distribution of LQ values in biotechnology at the level of NUTS2 regions (again measured relative to the EU-27 average). As the box plot graph shows, Germany has both the highest median value in terms of regional LQ values as well as particularly the highest LQ values for regions in the upper quartile of the distribution indicating that Germany hosts the most highly concentrated hot spots of European biotechnology employment. Looking at the absolute size of the biotechnology in the four economies (measured in terms of thousands of employees), panel (c) in Figure A.1 shows that biotechnology employment in the UK and Germany is above the employment level in France and Switzerland, particularly in recent years. Again the time series in panel (c) shows that France experienced a considerable drop in employment levels compared to the UK and Germany.

On the basis of different indicators (LQ, size and focus) the European cluster observatory also computes a 'star' ranking as composite indicator for the significance of the industry in a particular country or region (see Crawley and Pickernell, 2012, for a critical appraisal). The focus sub-indicator, which has not been presented here, thereby measures the extent to which the national economy is focused upon biotechnology employment relative to total national employment. The 'star' ranking in panel (d) of Figure A.1 indicates that Germany has outpaced the other three economies with respect to this composite ranking for biotechnology mainly during the recent sample period after 2000.

Further indicators for the evolution of the sectoral innovation system in German biotechnology point to the same direction: Comparing the distribution of patent applications in the EU-28 and the United States for the period 2004-2012, panel (a) of Figure A.2 shows that both economic blocks have an almost equal overall share (U.S. 52%; EU-28 48%). Within the EU-28, Germany has the largest share of patent applications followed by France and the UK. Looking at the evolution of patent applications over the time period 1977 to 2012, panel (b) of Figure A.2 shows that the number of EPO patent applications per million of inhabitants has grown exponentially during the 1990s in selected European economies and the United States. Thereby the German patent dynamics has even matched up with the level of patent applications per million inhabitants in the United States as leading biotechnology nation in the world. However, at the same time the evolution of patent applications since the second half of the 2000s shows a trend reversal, particularly for Germany and the United States. The figure thus underlines that particularly the 1990s and early 2000s mark a period of rising biotech activity in the United States and Europe.

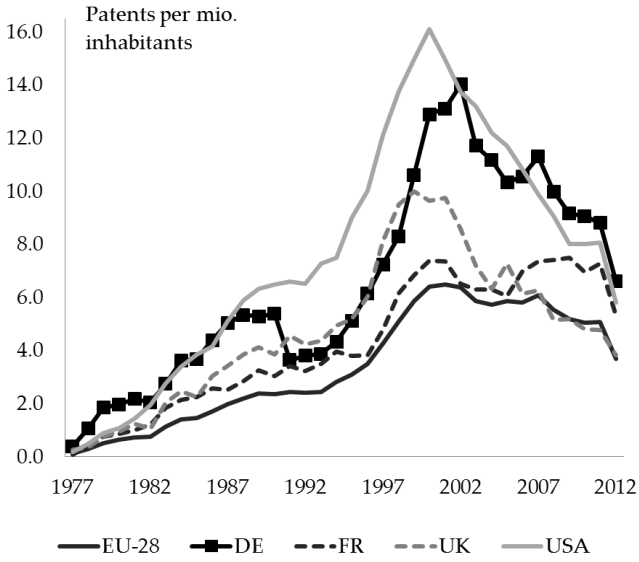
Figure A.2. Patent shares and evolution of patent applications in biotechnology per million inhabitants

(a) Patent Shares



Country	Patents in 2004-2012
Germany	7310.51
France	4026.02
UK	3136.54
Other EU-28	10166.4
USA	26257.89

(b) Patent Evolution



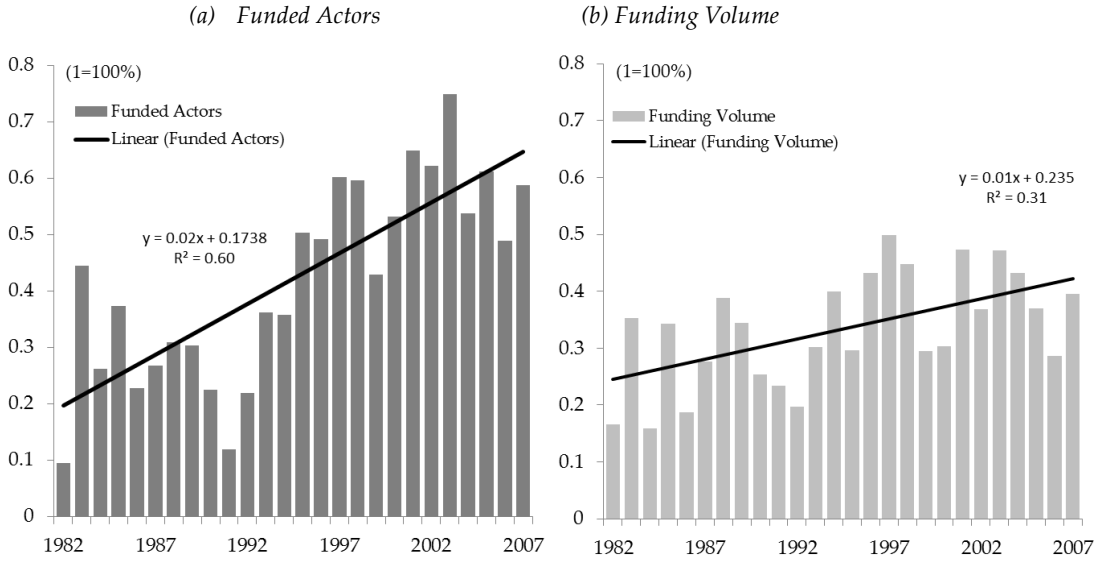
Notes: Panel (a) describes to percentage shares of patent applications for 2004-2012 in the EU-28 and the United States (summed to 100%); panel (b) show the time evolution for selected countries since 1977; both figures calculated on the basis of data from Eurostat (2016).

Taking a closer look at the German development during this period, particularly massive de-regulation, public funding, good provision of venture capital, a high rate of innovative start-ups and rapid localized knowledge transfers have been identified the major driving factors for the rapid progress of the industry in the 1990s (see, for instance, Dohse, 2000, Müller, 2002). The rise of the German biotechnology sector was thereby also supported by a global change in the worldwide technological regime of the industry, which evolved from an explorative state-of-art to a merely exploitative one. In the course of this structural shift, codified knowledge (rather than tacit) became increasingly important and facilitated the spreading of new ideas and research collaborations across longer distances (Ter Wal, 2014). These developments paved the way for the emergence of multiple hot-spots of biotech activity in the geographical landscape of the industry's innovation system.

One specific indicator for the change in the technological regime in biotechnology is the growing importance of collaborative R&D activities (see, for instance, Roijakkers and Hagedoorn, 2006, for the specific case of pharmaceutical biotechnology). Public support to collaborative R&D projects

thereby also turned into the main focus of policy makers. As Figure A.3 shows on the basis of data from the PROFI database of the German Federal Ministry for Education and Research (*Bundesministerium für Bildung und Forschung*, BMBF), the share of actors participating in funded collaborative R&D projects (panel (a)) as well as their relative funding volume as part of overall funding in the biotechnology sector (panel (b)) follow a positive long-run growth trend over the period 1982-2007. These growth trends thereby also reflect the paradigmatic shift in the organization of public funding schemes for the biotech industry since the mid-1990s on the basis of competitive funding elements and the promotion of cooperative R&D activities (BMBF, 2005).

Figure A.3. Share of actors and financial volume in publically-funded collaborate R&D projects



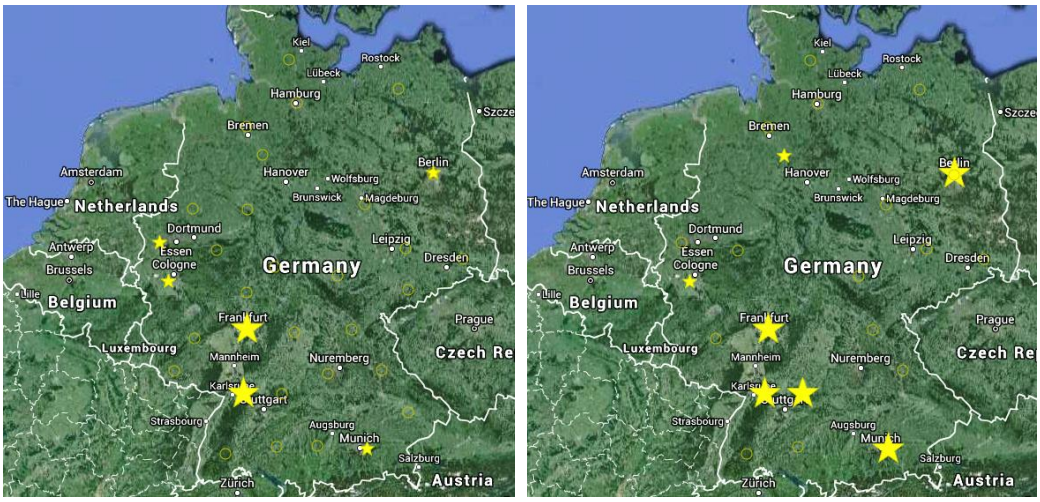
Notes: Panel (a) shows the percentage share of participating actors in collaborative R&D research projects in German biotechnology as part of all funded actors together with a linear trend line for period 1982-2007. Panel (b) shows the share of funding allocated to collaborative R&D research projects in German biotechnology as part of the overall funding volume together with a linear trend line for period 1982-2007. Calculations have been made on the basis of data from the PROFI database of the Federal Ministry of Research and Education. To identify relevant projects, the following technology fields have been selected: i) K - Biotechnology and ii) I19080 - Molecular Bioinformatics.

The *BioRegio* contest, launched in 1995, can thereby be seen as a forerunner for this new type of (cluster) policy schemes, which Eickelpasch and Fritsch (2005) label as 'contests-for-cooperation'. The main idea of the *BioRegio* contest was to encourage regional cluster initiatives (so-called *BioRe-*

gions) to form a common strategy and apply for subsidies to promote the biotechnology industry in the region (Dohse, 2000). A crucial feature of the *BioRegio* contest was its design as a network and cluster policy since the contest promoted the spatial clustering of biotechnology actors in regional innovation systems (Dohse, 2000). The underlying logic of this policy approach built on predictions from theoretical models in regional science and economic geography, which argued that spatial proximity and clustering of economic activities gives rise to increasing economies of scale through localization and urbanization advantages (McCann, 2013). The funding concept of the *BioRegio* contest thereby aimed at developing a new holistic approach for research and technology policy and was planned to integrate biotechnological capacities and scientific, economic and administrative activities.

Figure A.4. Spatial distribution of observatory star rating in biotechnology for German NUTS2 regions

(a) 1998 *(b) 2009*



NUTS2	Stars	LQ	Size	Focus	*BioRegion*
Karlsruhe	2	19.37	11.38	1.02	BioRegion Rhein-Neckar-Dreieck
Oberbayern	2	3.2	3.23	0.17	Initiativkreis Biotechnologie München
Berlin	2	2.56	1.66	0.13	BioTOP-Initiative Berlin-Brandenburg
Darmstadt	2	2.46	2.11	0.13	BioRegion Rhein-Main
Stuttgart	2	2.27	2.07	0.12	BioRegion Stuttgart/Neckar-Alb
Köln	1	1.75	1.47	0.09	BioRegion Rheinland
Niedersachsen	1	1.42	2.08	0.07	Region Nordwest-Niedersachsen

Notes: The table shows the values for sub-indicators used to calculate the cluster "stars" in 2009 together with a list of associated *BioRegions*; based on data from the European cluster observatory (2016).

Out of the 17 participating *BioRegions* an independent jury selected four winner regions (Rhineland, Rhine-Neckar, Munich and Jena) of the *BioRegio* contest. Selection criteria were mostly based on "hard" quantitative facts like the existence of a critical mass of biotech firms and research facilities within the region (for details, see Dohse, 2000). As Engel et al. (2013) point out, each winner region received a total lump sum amount of 25 million euro public grants (exception Jena: 15 million euro) for conducting joint R&D projects over a five-year time horizon (1997-2001). Additionally, the winner regions were favored in terms of getting access to the standard public funding schemes of the Federal Ministry of Research and Education. The total amount of these R&D grants exceeded 750 million euros for the time period 1997-2001. With regard to the share of publically funded collaborative R&D projects as outlined in Figure A.3, the four winner regions of the *BioRegio* contest received more than one third of the total collaborative R&D funding provided by the BMBF during the period 1997-2001. The *BioRegio* program was followed by smaller follow-up programmes such as *BioProfile* und *BioChance*.

Besides the purely monetary benefits for awarded *BioRegions*, participating in the contest was also considered attractive for non-winning participants, which could label themselves as part of the national network of *BioRegions* (organized as a registered association, for further information see *Arbeitskreis BioRegionen* https://www.biodeutschland.org/de/ak-bioregio.html) and potentially benefit from signaling effects due to the prestige of the contest. Moreover, the *BioRegio* contest was followed by the *BioProfile* contest starting in 1999 and its winners were mostly selected out of the original pool of *BioRegions*. As Figure A.4 visualizes for the "star" ranking from the European cluster observatory (based on German NUTS2 regions for the sample years 1998 and 2009), these funding programmes led to the prevalence of local clusters and fostered the development of strong regional nodes in the German biotechnology research network, which can be linked to associated *BioRegions* as shown in Figure A.4.

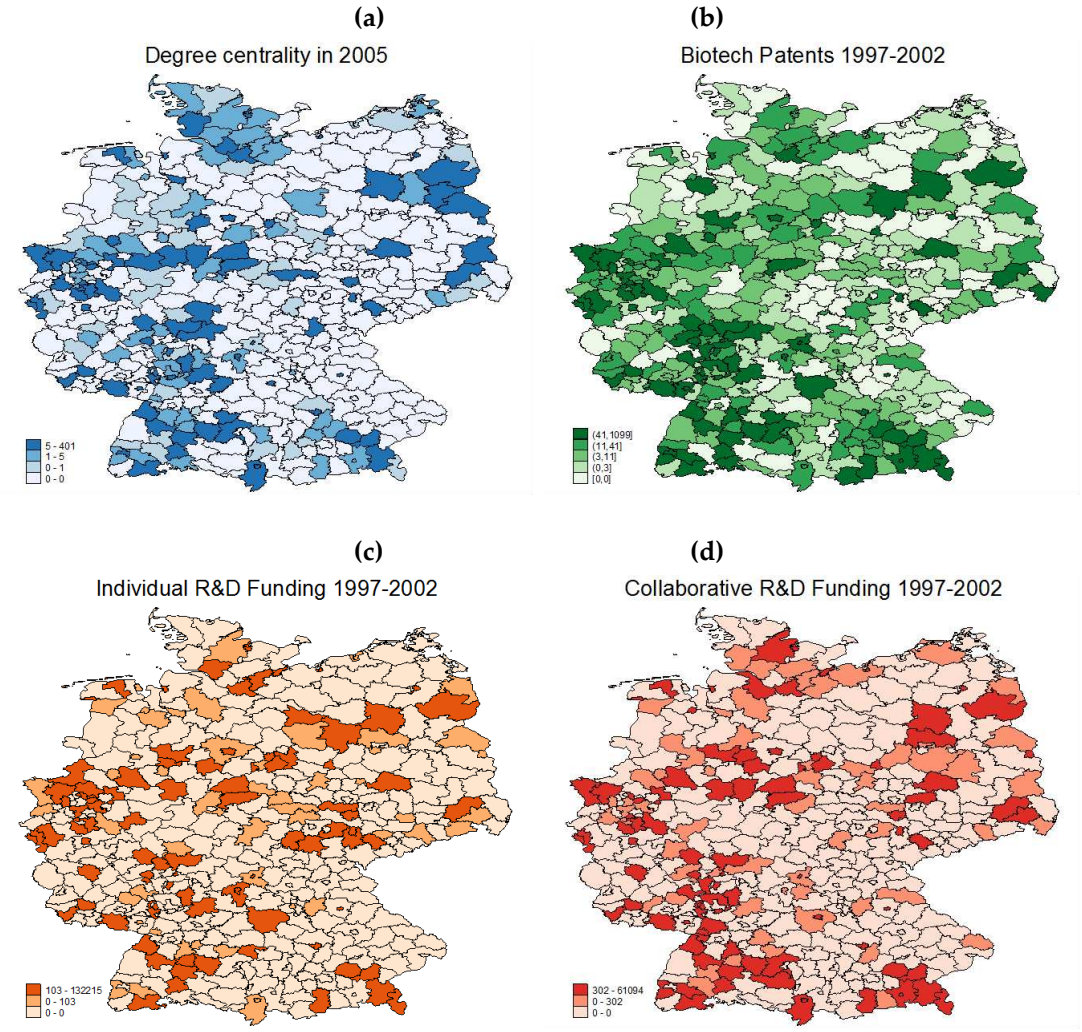
Appendix B: Summary statistics for variables and additional estimation results

Table B.1. Sample period and summary statistics of variables

Variable	Sample Period	Mean	Std. Dev.	Min.	Max.
Collaborative linkages	2005	0.014	0.406	0	102
Collaborative linkages	2009	0.013	0.363	0	87
Actors	2005	2.278	7.127	0	100
Loops	2005	0.708	4.166	0	102
Geographical distance	2005	308.74	152.74	0	844.50
Biotech patent applications	Sum of 1997-2002	16.058	46.893	0	576.06
Individual R&D funding	Sum of 1997-2002	1633.88	9728.61	0	1322145
Collaborative R&D funding	Sum of 1997-2002	1082.70	4645.32	0	61094
BioRegio dummy (winner x winner)	Binary dummy	0.001	0.035	0	1
BioRegio dummy (winner x participant)	Binary dummy	0.004	0.063	0	1
BioRegio dummy (participant x participant)	Binary dummy	0.004	0.060	0	1
High-tech start-ups	Average of 1996-2003	0.005	0.003	0	0.021
Medium-tech start-ups	Average of 1996-2003	0.007	0.003	0.001	0.036
KIS start-ups	Average of 1996-2003	0.063	0.020	0.021	0.154
International openness	Average of 1997-2002	26.050	13.572	0	96.186
MINT employment	Average of 1997-2002	2.198	1.290	0.450	13.550
Population density	Average of 1997-2002	514.27	662.51	40.84	3904.83
Sectoral specialization manufacturing	1998	713.35	901.53	144.48	8120.97
Sectoral specialization business-rel. services	1998	254.18	202.69	39.13	2911.67
Sectoral specialization household-rel. services	1998	129.69	98.59	19.53	609.07
Ellison-Glaeser index manufacturing	1998	27.741	35.697	0.210	390.83
Ellison-Glaeser index business-rel. services	1998	12.086	36.241	0.054	460.57
Ellison-Glaeser index household-rel. services	1998	6.753	23.263	0.013	279.04

Notes: See Table 4 in the main document for variable definitions.

Figure B.1. Spatial distribution of selected variables among German NUTS3 regions



Note: The four panels display the following information: (**a**) degree centrality as defined in eq.(1); (**b**) number of biotechnology patents (OECD definition); (**c**) volume of individual R&D funding and (**d**) volume of collaborative R&D funding. Patent applications and volumes of R&D funding are calculated as sum for the period 1997-2002 (see Table B.1).

Table B.2. Estimation results for the determinants of collaborative links among NUTS3 regions in 2009

Dep. Var.:	$collab_{ij}$ (in 2009)	**NegBin**	**ZINB**
	$collab_{ij}$ (in 2005)	0.426	0.349
		(0.055)***	(0.059)***
Core	Actors	0.044	0.051
		(0.007)***	(0.007)***
	Loops	-0.004	-0.003
		(0.003)	(0.003)
	Geographic distance	-0.681	-0.527
		(0.048)***	(0.068)***
	Biotech patent applications	0.254	-0.039
		(0.041)***	(0.104)
Policy	Individual R&D funding	0.126	0.091
		(0.015)***	(0.044)**
	Collaborative R&D funding	0.152	0.068
		(0.017)***	(0.061)
	BioRegio dummy (winner x winner)	-0.196	0.085
		(0.280)	(0.267)
	BioRegio dummy (winner x participant)	-0.069	0.047
		(0.193)	(0.188)
	BioRegio dummy (participant x participant)	0.258	0.291
		(0.216)	(0.213)
RIS	High-tech start-ups	0.480	0.491
		(0.180)***	(0.185)***
	Medium-tech start-ups	0.100	0.117
		(0.211)	(0.227)
	KIS start-ups	-0.137	-0.134
		(0.291)	(0.334)
	International openness	-0.183	-0.189
		(0.038)***	(0.040)***
	MINT employment	0.614	0.583
		(0.198)***	(0.202)***
	Population density	0.172	0.195
		(0.084)**	(0.087)**
	Specialization manufacturing	0.047	0.048
		(0.093)	(0.093)
	Specialization business-rel. services	-0.290	-0.227
		(0.128)**	(0.141)
	Specialization household-rel. services	0.167	0.151
		(0.090)*	(0.096)
	Ellison-Glaeser index manufacturing	-0.591	-0.547
		(0.094)***	(0.102)***
	Ellison-Glaeser index business-rel. services	0.836	0.838
		(0.279)***	(0.287)***
	Ellison-Glaeser index household-rel. services	-0.308	-0.326
		(0.220)	(0.225)
No. of observations		96,579	96,579
Wald test (χ^2) for spatial filters		180.81***	171.73***
LR-test (Poisson vs. NegBin)		180.16***	46.69***,$
Vuong (NegBin vs. ZINB)			1.13

Notes: * p<0.1; ** p<0.05; *** p<0.01; standard errors are given in brackets; $ = LR-test for zero-inflated Poisson vs. ZINB. See main document for details. A constant term has been included in the regression but is not reported here.

Table B.3. List of NUTS3 regions as a member of a *BioRegio* winner and participant cluster initiatives

ID	Name of NUTS3 district	*BioRegio* winner	*BioRegio* participant	*BioRegio* number
1002	Kiel	0	1	10
1003	Lübeck	0	1	10
2000	Hamburg	0	1	10
13003	Rostock	0	1	4
13001	Greifswald	0	1	4
3405	Wilhelmshaven	0	1	11
3403	Oldenburg	0	1	11
4011	Bremen	0	1	2
4012	Bremerhaven	0	1	2
3241	Region Hannover	0	1	2
3201	Hannover	0	1	9
3101	Braunschweig	0	1	9
3152	Göttingen	0	1	9
5124	Wuppertal	1	0	13
5111	Düsseldorf	1	0	13
5315	Köln	1	0	13
5313	Aachen	1	0	13
5316	Leverkusen	1	0	13
5354	Aachen	1	0	13
5358	Düren	1	0	13
5314	Bonn	1	0	13
6534	Marburg-Biedenkopf	0	1	7
6531	Gießen	0	1	7
6414	Wiesbaden	0	1	14
6412	Frankfurt	0	1	14
7315	Mainz	0	1	14
6411	Darmstadt	0	1	14
6413	Offenbach	0	1	14
6436	Main-Taunus	0	1	14
6438	Offenbach	0	1	14
7314	Ludwigshafen	1	0	15
7316	Neustadt a. d. W.	1	0	15
8111	Stuttgart	0	1	16
8116	Esslingen	0	1	16
8221	Heidelberg	1	0	15
8222	Mannheim	1	0	15
8416	Tübingen	0	1	16
8415	Reutlingen	0	1	16
8417	Zollernalbkreis	0	1	16
8311	Freiburg	0	1	3

Table B.3. (cont'd) List of NUTS3 regions as a member of a *BioRegio* winner and participant cluster initiatives

ID	Name of NUTS3 district	*BioRegio* winner	*BioRegio* participant	*BioRegion* number
8421	Ulm	0	1	17
9162	München	1	0	8
9188	Starnberg	1	0	8
9362	Regensburg	0	1	12
16053	Jena	1	0	6
15202	Halle	0	1	5
14365	Leipzig	0	1	5
15261	Merseburg-Querfurt	0	1	5
15265	Saalkreis	0	1	5
15154	Bitterfeld	0	1	5
11000	Berlin	0	1	1
12065	Oberhavel	0	1	1
12069	Potsdam-Mittelmark	0	1	1
12072	Teltow-Fläming	0	1	1
12054	Potsdam	0	1	1

Note: The *BioRegion* number refers to the definition given in Figure 2 in the main document.

Additional References (for Appendix A and B)

BioSpace (2009): France biotech reports a 79% fall in equity investments in French biotech companies in 2008 and asks the government to implement a stimulus plan for young, innovative companies. Available at (last accessed: 05.07.2016): http://www.biospace.com/News/-france-biotech-reports-a-79-fall-in-equity/125381.

BMBF (2005): BioRegionen in Deutschland. Starke Impulse für die nationale Technologieentwicklung. 3rd edition, Bundesministerium für Bildung und Forschung (BMBF): Bonn und Berlin.

Crawley, A.; Pickernell, D. (2012): An appraisal of the European Cluster Observatory, in: European Urban and Regional Studies, 19(2): 207-211.

Dickman, S. (1996): Germany joins the biotech race, in: Science, 274: 1454-1455.

Dohse, D. (2000): Technology policy and the regions – the case of the BioRegio contest, in: Research Policy, 29: 1111-1133.

Eickelpasch, A.; Fritsch, M. (2005): Contests for cooperation – A new approach in German innovation policy, in: Research Policy, 34(8): 1269-1282.

European cluster observatory (2016): Clusters at your fingertips, available at (last accessed 05.07.2016): http://www.clusterobservatory.eu/.

Eurostat (2016): Biotechnology patent applications to the EPO by priority year, available at (last accessed 18.01.2017): http://ec.europa.eu/eurostat/en/web/products-datasets/-/PAT_EP_NBIO

France Biotech (2008): The France biotech annual biotechnology industry report 2007 / 2008. Available at (last accessed 05.07.2016): www.france-biotech.org/en.

Isserman, A. (1977): The location quotient approach to estimating regional economic impacts, in: Journal of the American Institute of Planners, 43(1): 33-41.

Laskawi, C. (2015): Biotechnology. Funding gap jeopardizing competitiveness. Current Issues, Deutsche Bank Research, available at (last accessed 04.07.2016): www.dbresearch.com.

Müller, C. (2002): The evolution of the biotechnology industry in Germany, in: Trends in Biotechnology, 20(7): 287–290.

PharmaLetter (2005a): French biotechnology sector investment is waning, says France Biotech. Available at (last accessed 05.07.2016): http://www.thepharmaletter.com/article/-french-biotechnology-sector-investment-is-waning-says-france-biotech.

PharmaLetter (2005b): French biotechnology confronts financial crisis in the industry. Available at (last accessed 05.07.2016): http://www.thepharmaletter.com/article/french-biotechnology-confronts-financial-crisis-in-the-industry.

Roijakkers, N.; Hagedoorn, J. (2006): Inter-firm R&D partnering in pharmaceutical biotechnology since 1975: Trends, patterns, and networks, in: Research Policy, 35(3): 431-446.