RUHR
ECONOMIC PAPERS

Timo Mitze
Falk Strotebeck

Parameters of Regional Cooperative Behavior in the German Biotech Industry

A Quantitative Social Network Analysis

RUHR UNIVERSITÄT BOCHUM RUB

#374

Imprint

Ruhr Economic Papers

Published by

Ruhr-Universität Bochum (RUB), Department of Economics
Universitätsstr. 150, 44801 Bochum, Germany

Technische Universität Dortmund, Department of Economic and Social Sciences
Vogelpothsweg 87, 44227 Dortmund, Germany

Universität Duisburg-Essen, Department of Economics
Universitätsstr. 12, 45117 Essen, Germany

Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI)
Hohenzollernstr. 1-3, 45128 Essen, Germany

Editors

Prof. Dr. Thomas K. Bauer
RUB, Department of Economics, Empirical Economics
Phone: +49 (0) 234/3 22 83 41, e-mail: thomas.bauer@rub.de

Prof. Dr. Wolfgang Leininger
Technische Universität Dortmund, Department of Economic and Social Sciences
Economics – Microeconomics
Phone: +49 (0) 231/7 55-3297, email: W.Leininger@wiso.uni-dortmund.de

Prof. Dr. Volker Clausen
University of Duisburg-Essen, Department of Economics
International Economics
Phone: +49 (0) 201/1 83-3655, e-mail: vclausen@vwl.uni-due.de

Prof. Dr. Christoph M. Schmidt
RWI, Phone: +49 (0) 201/81 49-227, e-mail: christoph.schmidt@rwi-essen.de

Editorial Office

Joachim Schmidt
RWI, Phone: +49 (0) 201/81 49-292, e-mail: joachim.schmidt@rwi-essen.de

Ruhr Economic Papers #374

Responsible Editor: Thomas K. Bauer

ISSN 1864-4872 (online) – ISBN 978-3-86788-429-7

Ruhr Economic Papers #374

Timo Mitze and Falk Strotebeck

Parameters of Regional Cooperative Behavior in the German Biotech Industry

A Quantitative Social Network Analysis

RUHR UNIVERSITÄT BOCHUM RUB

Bibliografische Informationen der Deutschen Nationalbibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über: *http://dnb.d-nb.de* abrufbar.

http://dx.doi.org/10.4419/86788429
ISSN 1864-4872 (online)
ISBN 978-3-86788-429-7

Timo Mitze and Falk Strotebeck[1]

Parameters of Regional Cooperative Behavior in the German Biotech Industry – A Quantitative Social Network Analysis

Abstract

We analyse the determinants of network formation in Germany's biotechnology industry using social network analysis combined with a regression approach for count data. Outcome variable of interest is the degree centrality of German regions, which is specified as a function of the region's innovative and economic performance as well as biotech-related policy variables. The inclusion of the latter allows us to shed new light on the question to what extent R&D-based cluster policies are able to impact on the formation of the German biotech network. Our results show that policy indicators such as the volume of public funding for collaborative R&D activity are positively correlated with the region's overall and interregional degree centrality. However, besides this direct funding effect, we do not observe any further (non-pecuniary) advantages such as prestige or image effects. Regarding the role played by locational factors as elements of the sector-specific and broader regional innovation system, we find that the number of biotech patent applications, the share of regional hightech start-ups and the population density among other factors are positively correlated with the region's position in the German biotechnology network.

JEL Classification: C21, R38

Keywords: Biotechnology; network formation; degree centrality; R&D policy

October 2012

1 Timo Mitze, University of Southern Denmark; Falk Strotebeck, Ruhr-Universität Bochum. – *The authors thank Ann-Katrin Voit for helpful comments. –* *All correspondence to Timo Mitze, Department of Border Region Studies, University of Southern Denmark, Alsion 2, 6400 Sønderborg, Denmark, E-Mail: tmitze@sam.sdu.dk.*

1. Introduction

Economic systems are characterized by mutual interdependencies among their actors and the emergence of network structures is a crucial pillar of the knowledge economy. For instance, having access to distinct knowledge networks is an important prerequisite to build up individual stocks of knowledge and benefit from effective knowledge transfer among actors. Empirical research on network complexity and dynamics is therefore a challenging perspective (Schweitzer et al., 2009).

In this paper, we seek to analyse the network structure of the German biotechnology industry and its determinants. Given that the contemporary regional distribution of this industry shows a particular spatial structure with the evolution of distinct regional clusters (Ter Wal, 2009), we link the aggregate industrial analysis with a particular regional perspective. Our motivation for doing so basically stems from two facts: On the one hand, assessing the national and local economic performance of the biotech sector and tracking its industrial evolution has started to become an attractive field of research for different disciplines including economics and business sciences as well as geographers and regional scientists. On the other hand, as Henkel and Maurer (2010) argue, the analysis of network effects in biotechnology is still relatively overlooked compared to other hightech industries such as software and electronics.

For our empirical investigation we use a unique dataset, which combines firm level information with sectoral and regional indicators, and apply social network analysis (SNA) as well as a quantitative regression approach. Units of our analysis are German biotech regions at the NUTS3 level. In a first step, SNA is used to quantify and visualize flows of knowledge among the German biotech industry and compute key characteristics of the underlying regional and interregional networks. Though SNA is a powerful tool for this type of analysis (see Ter Wal and Boschma, 2009), the approach is often critized for lacking the methodological depth to conduct a thorough identification strategy of the key motivations behind certain structural relationships. To overcome this shortcoming, we use the results obtained from SNA (i.e. the region's degree centrality) as outcome variable in a quantitative regression approach to gain further insights into the link between regional economic factors and the region's network position.

Throughout the empirical analysis we put a particular focus on the analysis of the effectiveness of policy instruments and their impact on the individual region's knowledge ties. Since public R&D support to selected regional biotech clusters (so called *BioRegios*) has a high priority for the Federal Government to increase the international competitiveness of the German biotech industry, stimulating R&D cooperations is an

important intermediate target of R&D policy with the aim to strengthen the international competitiveness of the industry and its individual actors in the long-run. By focusing on this intermediate target as a necessary prerequisite for policy success, our analysis may be seen as a complementary attempt to evaluate the success of R&D-based cluster policies on top of the assessment of standard outcome variables such as the regional patent activity (see e.g. Engel et al., 2012, for a recent application).

Foreshadowing some empirical results, the network analysis shows that the structure of the biotechnology network is far from being random and features specific industrial and regional characteristics. The SNA identifies a core network comprising particular drivers of German biotechnology. Although cooperation in general is highly localized, cooperative linkages between regions over long distances mark further important channels of knowledge flows and can be seen as crucial bridges between local networks.[2] In the regression approach we additionally show that - in order to benefit from these interregional bridges for a diffusion of knowledge in the sector - regions need to have a sufficiently high absorptive capacity. From a sectoral perspective, the number of biotech patent applications is positively correlated with the number of realized cooperation linkages. From a broader regional perspective, also further modern location factors such as regional labor market conditions, the entrepreneurial spirit and general agglomeration forces show to matter as locational factors attracting cooperative activity.

Regarding the role of R&D policy, clearly, the amount of cooperative R&D funding allocated to each region drives its position in the biotech network. However, compared to cooperative funding we do not observe any significant role played by individual R&D support. Moreover, opposed to the direct funding effect from the provision of collaborative R&D grants, we also do not observe any additional indirect impacts such as mobilization and prestige effects etc. on the degree centrality of those regions that were either a winner or participant of the BioRegio and BioProfile contests of cooperation.

The remainder of the paper is organized as follows: The next section gives a description of the structure of regional networks in the German biotech industry by means of SNA. We thereby focus on different levels of regional aggregation. Section 3 reviews the theoretical and empirical literature dealing with network determinants to work out an empirical model of regional network formation. We put a special focus on policy measures and factors driving the regional innovation system. Having derived testable hypotheses, Section 4 then presents the empirical results of our estimation approach

[2] 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 biotechnology industry (see, e.g. Gertler and Levitte, 2005).

using data for the region's total degree centrality with respect to all other regions in the sample as well as dyadic data which explicitly models individual linkages between each pair of regions and accounts for heterogeneity among each 2-tuple of regions. Section 5 concludes the paper and gives some policy recommendations.

2. The Structure of Regional Networks in the German Biotechnology Industry

Started as an economy with a merely inhospitable climate for biotechnology two decades ago (Dickman, 1996), today Germany is one of the leading European biotech countries. Massive public funding, good provision of venture capital, a high rate of innovative start-ups and rapid localized knowledge transfer are typically seen as the major driving factors for this progress. The rise of the German biotechnology sector was also supported by a global change in the worldwide vision 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 cooperations across longer distances (Ter Wal, 2009). These developments paved the way for the emergence of multiple hot-spots of biotech activity in the geographical landscape of Germany.

We will take these stylized facts as a starting point for our SNA. Basis of the empirical analysis is a firm level dataset from the BIOCOM Year- and Addressbook 2005, which contains basic information on 1709 firms and institutions of the biotech sector as well as their cooperation activities.[3] The actual kind of cooperation is unfortunately not precisely specified in the dataset. Therefore, internet research has been conducted to exclude pure supplier relationships (objects of utility like petri dishes or pipettes) and advisory services (such as consultancies). This reduction leads to a set of 575 core-firms of the German biotech sector with R&D as main business activity and a strong focus on R&D cooperations. 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) cooperation partner not included in the set of core-firms was added to the latter. This finally results in a dataset of 1002 firms and institutions, which will be used in the following.

Since we are especially interested in the regional dimension of the biotech network, firms and institutions have been aggregated to the district level (NUTS3). Out of the 439

[3] A verification rate of 84 percent of the whole dataset is declared by the BIOCOM AG after the actualization period in 2004. Nevertheless it has to be mentioned, that the information regarding the characteristics of the firms is given on a voluntarily basis (see BIOCOM AG, 2005).

German districts, 178 districts contain firms and/or institutions relevant for our analysis. In a further step, these 178 districts have been mapped into a set of *BioRegios*. These macro-regions have been formed in the course of the BioRegio contest in the mid-1990s as a cluster-based development strategy. In total, 17 *BioRegio* clusters applied for R&D funding, out of which 4 clusters have been awarded as "winner" by an independent jury. Selection criteria were mostly based on "hard" facts like the existence of a critical mass of biotech firms and research facilities within the region (for details, see Dohse, 2000). Each winner received a fixed amount of public grants and was favored in terms of getting access to the standard R&D-grant schemes over the period of 1997 to 2001. Next to these pecuniary incentives, being a *BioRegio* was not only attractive for winners: Also non-winning participants could label themselves as part of the national network of *BioRegios*, which was organized as a registered association, and potentially benefit from its prestige. Moreover, the BioRegio competition was followed by the BioProfile contest starting in 1999 and its winners were mostly selected out of the original pool of *BioRegios*.[4]

Given that the initial 17 *BioRegios* are a crucial backbone of the German biotechnology industry, in Figure 1 we look at the district-to-district cooperation network grouped by these *BioRegios*. Each vertex presents one *BioRegio* and edges of the network visualize cooperative linkages. We count all district-to-district linkages of firms within a *BioRegio* as loops. In this way we can easily distinguish among cooperations that are intraregional and interregional in nature.[5] Given these definitions, the graphical interpretation of Figure 1 is rather straightforward: The absolute size of the vertices indicates the number of loops within each *BioRegio*, whereas the width of each edge combining two vertices highlights the number of cooperation linkages observed between the macro regions.

<< Figure 1 about here >>

As Figure 1 shows, the German biotech network is characterized by mutual interdependencies among the different *BioRegios*. We also observe a high degree of regional heterogeneity, both with respect to intraregional as well as interregional cooperations. The rather big "99"-vertex of regional actors outside any established *BioRegio* demonstrates that German biotech activity is not solely focused within these regional clusters. With 378 actors within the remainder "99"-vertex nearly 37 percent of

[4] A detailed list of regions participating in both contests is given in Table A.1 in the appendix.
[5] Obviously, the number of loops increases with a higher aggregation level. Moreover, the number of cooperation linkages on interregional level also depends on the number of biotech firms within a district or biotech cluster.

actors from our dataset are not part of a *Bioregio*; however, this also means that 63 percent indeed are. Among the *BioRegio* clusters the biggest player characterized both by vertex size and width of edges is the BioTOP Initiative Berlin-Brandenburg with 133 actors. As Table 1 shows, the remaining *BioRegios* consist – on average – of about 30 actors.

<< Table 1 about here >>

Based on this basic information, we can compute two indicators relevant for the SNA: 1.) the *degree centrality* (C^D) of a vertex (v_k) counts the number of direct linkages of an actor and sample size *N*, where the function $f(v_i, v_k)$ is one, when v_k and v_i are directly linked and zero otherwise (Freeman, 1978/79)

(1) $$C^D(v_k) = \sum_{i=1}^{N} f(v_i, v_k)$$

2.) the *average degree* then is just the sum of all degree centrality values (of all v_i) divided by the number of actors. As shown for the case of the BioTOP Initiative Berlin-Brandenburg, the number of loops and interregional linkages are likely to increase with the number of firms and institutions in a particular region. However, this relationship does not appear to be a linear one: For example, the BioRegio 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 (2.39 compared to 3.39, respectively). This interregional heterogeneity becomes also visible if we compare the BioRegio Rhein-Main (No. 14) with BioRegio Stuttgart/Neckar-Alb (No. 16): While the Rhein-Main region contains more actors (firms/institutions) than the cluster in 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 degree centrality to the number of its actors.

To further investigate this issue, we take a closer look on how cooperative linkages (edges) are distributed between the *BioRegios*. Table 2 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 (internal and external) cooperative linkages with Bremen (No. 2), while the relative importance for cooperations with Berlin-Brandenburg from the perspective of Bremen is only 3 percent. Thus, the relative importance of interregional cooperation activity is not symmetrical for the different regions. From the results in Table 2 one can deduce the relative importance of the set of *BioRegios*.

<< Table 2 about here >>

Another interesting result from Table 2 is shown if we look at entries on the diagonal (bold values), which display the percentage of cooperation linkages defined as loops. Here again, there are strong regional differences among the *BioRegio*s. For instance, the regional actors of the BioRegio 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 *BioRegio* ("99"-vertex) and roughly 17 percent between Jena and the BioRegio Halle-Leipzig. This indicates that the role of interregional linkages may be important, especially for small *BioRegios*. In comparison to Jena, the share of intraregional cooperations for the large BioTOP Initiative Berlin-Brandenburg is 25 percentage points higher (in total 38 percent). This large difference highlights that the Berlin-Brandenburg region has a sufficiently high internal absorptive capacity for R&D cooperations, while the BioRegio Jena heavily depends on external research partners.

Looking at the geographical distribution of interregional cooperations, a striking fact is that the actors in Jena tend to cooperate with external partners in close geographical proximity (Halle-Leipzig). Although Halle-Leipzig is a rather small regional cluster (28 actors), the relative share of cooperations from regional actors in Jena is much higher compared to relatively large clusters such as Berlin-Brandenburg (133 actors, 10 percent) as well as BioRegio Rheinland (74 actors, no cooperation). Thus, geographical proximity seems to matter in determining cooperative linkages among the *BioRegios*. The explicit role of distance is also underlined by the fact that in nearly all cases, the highest weight is given to internal cooperations (values on the diagonal), even though there is a remarkable difference among clusters, ranging from a low value of 10 percent (Initiativkreis Biotechnologie München, No. 8) to a high share of 31 percent (BioRegio Greifswald-Rostock, No. 4) or even 38 percent (BioTOP-Initiative Berlin-Brandenburg).

All in all, our SNA results show that cooperative behavior varies considerably by *BioRegios*. This heterogeneity can also be seen if we additionally plot the distribution of the average degree at the NUTS3 level. As Figure 2 shows, there is a wide range of observable values: In 68 out of the 178 districts the average degree of an actor is between zero and one, while few actors even show values up to nine. This regional diversity can finally also be gathered if we draw a map of the biotech network at the district level in Figure 3. The figure shows that large agglomerations such as Berlin (11000), Hamburg (02000) and Munich (09162, 09184) are at the core of the network.

However, also a wide range of smaller districts such as Heidelberg (08221), Tübingen (08461) and Freiburg (08311) show a strong performance in terms of cooperative behavior. Since all of the latter actors are embedded in larger *BioRegio* clusters, this advocates the need to elaborate an exploratory concept of network formation accounting for the role played by the sectoral and regional innovation system and test for its explanatory power in an empirical regression exercise. We will do so in the next section.

<< Figure 2 about here >>

<< Figure 3 about here >>

3. Network Determinants: Regional Innovation System and Policy Measures

As Cassi and Plunket (2010) point out, "*the dynamics of network formation remains an under-investigated question*". To fill this gap, the authors propose a categorization of network determinants based on two concepts: 1.) the relational perspective, and 2.) proximity mechanisms. While the first approach deals with issues such as 'closure', indicating that local networks become more dense and clustered, as well as 'preferential attachment', where a new actor will most likely attach to an actor that already has a high number of linkages, the second concept mainly refers to different measures of proximity as drivers of network formation, including geographical proximity but also alternative kinds like cognitive, organizational, social and institutional distance (see Boschma, 2005). In the following, we set up an eclectic modelling framework that will take up elements from both concepts.

At the heart of the relational perspective is the standard preferential attachment model (Barabási and Albert, 1999), which starts from the simple logic, namely that the probability for initiating new cooperations is a function of the stock of realized ones. Since the standard model solely builds upon path dependency as driver of network cooperations, recent extension as in Bianconi and Barabási (2001) have attempted to make the concept more suitable for structural analyses: This is, for instance, done by adding an endogenous fitness parameter to the model, which impacts on the likelihood of being chosen as a potential cooperation partner. The resulting 'fitter-get-richer' model thus allows the formation of an explanatory theory of network formation, which is not solely a function of path dependency based on historical stocks of cooperations but instead also allows initially less connected actors to catch up or even get ahead of

actors with an initially higher stock of cooperations. Barabási (2003) describes this fitness parameter as "*a quantitative measure of a node's ability to stay in front of the competition*."

We link the basic idea of the 'fitter-get-richer' approach to the notion of regional competitiveness, which is in the focus of the research literature on regional innovation systems. The latter field gives a systematic account of the different factors driving regional competitiveness. Timmeren and Röling (2007), for instance, apply the 'fitter-gets-richer' model to urban development. Indeed, locational factors appear to be quite relevant for the formation of network structures, as Glückler (2007) points out that *"[…] place makes a difference. […] a place may be conceived as a bundle of resources and opportunities with the additional characteristic of spatial contiguity. […] This localized resource profile comprises the structural aspects of relationships (e.g. social capital, structural holes) as well as the material, social and institutional resources the these relationships access and transfer*." This argument is pretty much in line with what Marshall (1895) has already described as the benefits from agglomeration when thinking about the role of places. In his work, Marshall stresses the importance of technological spillovers, the existence of a pool of specialized workers, services and input providers, enhanced by a fourth factor, the size of the market (Echeverri-Carroll, Brennan 1999; Krugman 1991).

The innovation system approach takes up these elements and puts a special focus on social interaction in the production, diffusion and use of new knowledge (Lundvall, 1992, Edquist, 2000). Key indicators for the analysis of national and regional systems of innovation are the internal organisation of firms and interfirm relationships, the industry specialization and structure, the role of the public sector, the R&D intensity as well as the qualification and training system. According to Howells (1999), regional innovation systems thus represent crucial spots for localized learning and tacit know-how sharing. In extension to standard production location factors in the Marshallian concept, Falck and Heblich (2008) label these additional parameters as modern location factors based on inter-industry linkages, knowledge based inputs and creative entrepreneurs. As these points make clear, internal and external linkages are among the key determinants of a regional innovation system. At the same time, the ability of a region to be a vehicle for successful interactions is strongly influenced by its industrial, institutional and educational capacities. This latter link will be more closely analysed in the following.

Among the parameters that are likely to influence the cooperation intensity of regions, in first place we look for "hard facts" such as regional specialization with a critical mass in biotech activities. Clearly, the number of biotech firms in the region and regional R&D output measured as biotech-related patents are expected to positively influence the absorptive capacity of regions to engage in regional and interregional cooperation in the German biotech network. From a broader regional perspective, also the regional labor market conditions, the region's human capital endowment and entrepreneurial spirit as well as general localization and urbanization forces are likely to impact on the regions competitive capacity. To test for the different channels through which the regional innovation system can impact on the region's network position, we use a broad set of regional control variables in the regression approach. A full list of the set of regressors together with descriptive statistics is given in Table 3.

From the perspective of proximity mechanisms, an additional factor that is likely to influence the connectivity of regions is the geographical distance between actors/regions. Network relationships are supposed to be more common over short than long distances (Maggioni et al., 2007). As the SNA has already shown, the regional intensity of intraregional cooperations in *BioRegio* clusters is much higher compared to interregional cooperations. Moreover, among interregional cooperations, regional actors tend to choose external cooperation partners nearby (as shown for the example of the BioRegio Jena). However, as Ter Wal (2009) reviews recent findings from other network mapping approaches, networks are neither spread homogeneously across clusters nor are they confined to precise regional boundaries. For the set of *BioRegios* we might as well expect a tendency for 'closure', in other words, for building up linkages with near partners and exclude external partners from the sub-network.

Finally, policy variables may have an impact on the region's fitness parameter: Looking at the biotech industry in Germany, public funding has a prominent role and the BioRegio contest in the mid-1990s can been seen as a "kick-off" event for massive policy support. The underlying intention of the BioRegio contest and its successors was to support the regions with the best chances of success (Engel, Heneric 2005; Dohse 2005). Derived from this, one can assume that there has been a network of localized cooperating actors at the time the regions were competing for support. The further evolution of the network might have been influenced in different ways: According to the concept of preferential attachment, new nodes (actors like new firms and/or institutions) might have focused on building up cooperation linkages with actors, already in central positions within the network or especially successful.

If a *Bioregio* gains support, the region's degree centrality is likely to increase for two reasons. Firstly, there can be the direct effect mentioned before: The bioregion already have had a noteworthy cooperation basis when started competing for financial support. And secondly, a winning region might be seen as a pool of worthwhile partners to cooperate with. Regarding the latter effect, being a winner (having high chances to succeed in the biotech sector) might work as a signal for other actors in the biotech industry. However, the local network can show the tendency for closure and one may expect that, while intraregional linkages increase in the course of funding, interregional cooperative linkages are less likely to occur.

4. A Count Data Approach to Identify Factors Driving the Network Position

In this section we embed the SNA in a statistical modelling approach to formally test the relative importance of determinants for the region's degree centrality. As Knoke and Yang (2008) point out, besides a visual inspection of the data and the computation of summary measures, recently different tools have been developed that allow a deeper examination of the main parameters driving (dyadic) network ties as a function of the individual actors and aggregated explanatory factors. A prominent example is the p^* model, which applies a logistic regression approach to social network data with dichotomous directed ties among a set of g actors (Wasserman and Pattison, 1996).

We follow a similar quantitative approach here. However, since we are using regional data for $i=1,...,N$ regions, each aggregated over $r \in g$ actors, our data set is composed of count data rather than binary entries in the network matrix as typically used in the p^* approach. We thus apply a generalized linear regression approach based on the degree centrality as outcome variable in the model. Since we are interested in the determinants of cooperation intensity (rather the absolute number of cooperative links), we need to normalize the outcome variable as typically done in the field of SNA. Rather than calculating a measure for the average degree prior to estimation, we add the total number of biotech firms per NUTS3 region as additional right hand side control variable in the regression approach and are thus able to use the degree centrality as defined in eq.(1) as dependent variable in the regression equation.

In a first set of estimations, we use a model specification, which computes the total degree centrality for each of the N=439 NUTS3 regions in the dataset as the sum of linkages between region i and *all other regions*. As alternative estimation strategy, we also use a dyadic approach that models the pairwise number of cooperative linkages for

each (i, j)-region tuple in the 439x439 link matrix.[6] Since this matrix has symmetric entries, for estimation purposes we only rely on those observations in the lower triangular part in order to avoid overfitting the model by using double counts. Thus, for the disaggregate regression specification, this leaves us with a total of (440x439)/2=96580 observations on interregional cooperations of biotech firms for each NUTS3 region. One advantage of the disaggregate specification is that we have a higher number of observations at hand, which increases estimation efficiency. Additionally, the pairwise estimation approach for each (i, j)-region tuple allows to explicitly test for the influence of geographic distance between NUTS3 regions as typically assumed in the literature (see, e.g., Ter Wal, 2009).

In both the overall and pairwise specification we have to deal with a large number of zero observations since only 178 of all 439 NUTS3 districts host at least one biotech actor $(g \geq 1)$ registered in the BIOCOM database. Thus, standard Poisson or negative binominal regression models for count data may be biased by this inflation of zero values. A solution to this problem is to rely on zero-inflated Poisson or negative binominal specifications. Zero-inflated models generally assume different data generating processes to be in order when predicting the probability for having any collaboration, on the one hand, and its actual (non-zero) size, on the other hand. The first part of the model is estimated in terms of a binary choice model (logit or probit), which is then mapped into a standard Poisson or negative binominal specification. To guide model selection, different statistical tests will be used: To judge whether a standard Poisson distribution with equal mean and variance is valid compared to a negative binominal model with under- or overdispersion in this relationship, we use a standard Likelihood Ratio test for the statistical significance of the overdispersion parameter in the empirical model. Likewise, a Vuong (1989) test will be applied in order to discriminate between the standard and zero-inflated specification.

In the most general case of the zero-inflated negative binominal model (ZINB), we start from a negative binominal distribution for a variable Y as

$$(2) \qquad P(Y = y) = \frac{\Gamma(y+\tau)}{y!\Gamma(\tau)}\left(\frac{\tau}{\lambda+\tau}\right)^{\tau}\left(\frac{\lambda}{\lambda+\tau}\right)^{y}, y = 0,1,\ldots : \lambda, \tau > 0,$$

where $\lambda = E(Y)$ denotes the mean, τ is the shape parameter quantifying the amount of overdispersion. The variance for variable Y is defined as $\left(\lambda + \lambda^2 / \tau\right)$. For large values of

[6] Where the number of regions is $i, j = (1,\ldots,439)$ for the total degree centrality and $i, j = (1,\ldots,439 \mid j \neq i)$ for the subset of interregional linkages excluding loops.

τ, the negative binominal distribution approach to a Poisson distribution with $E(Y) = Var(Y) = \lambda$. From eq.(2), the ZINB distribution follows as

$$P(Y=y)\begin{cases} p + (1-p)\left(1+\frac{\lambda}{\tau}\right)^{-\tau}, y=0 \\ (1-p)\left(\frac{\Gamma(y+\tau)}{y!\Gamma(\tau)}\right)\left(1+\frac{\lambda}{\tau}\right)^{-\tau}\left(1+\frac{\tau}{\lambda}\right)^{-y}, y=1,2,\ldots, \end{cases} \quad (3)$$

where p is the proportion of extra zeros in the distribution of Y and $(1-p)$ is the proportion of non-zero values according to the negative binominal distribution. The mean and variance of the ZINB are $E(Y) = (1-p)\lambda$ and $Var(Y) = (1-p)\lambda(1+p\lambda+\lambda/\tau)$. For large τ the ZINB reduces to a zero-inflated Poisson distribution and for values of p close to zero, eq.(2) reduces to a standard negative binominal distribution.

To estimate the degree centrality (C^D) for region *i*, the ZINB based regression model can then be stated as

$$\lambda_i = \exp(\beta' \mathbf{x}_i + u_i) \text{ and } Prob(p_i) = \gamma' \mathbf{z}_i + \varepsilon_i, \quad (4)$$

where the non-zero information in C^D by means of λ_i is related to a vector of explanatory variables $\mathbf{x}_i$, and in the binary choice part of the model (here: probit specification) p_i measures the probability that C^D has zero entries, which is related to a vector of explanatory variables $\mathbf{z}_i$.[7] Further, u_i and ε_i are the residuals in the negative binominal and probit part, respectively, with $\exp(u_i) = Gamma(1/\tau, \tau)$ and $\varepsilon_i = N(0, \sigma^2)$. Empirical estimation of the model in eq.(4) is done by means of Maximum Likelihood (ML) techniques.[8] For the pairwise estimation model, the index *i* is substituted by *ij*, which increases the number of observations as outlined above.

For the estimation of eq. (4), we use a broad set of regional determinants as explanatory variables in $\mathbf{x}$ and $\mathbf{z}$. In order to account for the likely problem of reversed causality between degree centrality as outcome variable and regional (biotech related) factors, we put a specific time lag structure on the model. That is, we use the actual observation in 2005 for our dependent variable as well as for the number of biotech firms as normalizing factor. However, for all other time-varying explanatory variables we assume a time lag of at least three years. Although this is not a perfect strategy to reduce the risk of reverse causality, it allows us to interpret the regression results (carefully) as causal

[7] We allow for the case that $\mathbf{x}_i$ and $\mathbf{z}_i$ may potentially contain the same set of variables.
[8] For a formulation of the (log) likelihood function of the ZINB see, for instance, Mwalili et al. (2008). For estimation we use the *zinb* command written in Stata.

impacts on the region's degree centrality stemming from factors of the regional innovation system and the policy context.[9]

As example to motivate the need of imposing a lag structure in the model, we take a closer look at the link between public R&D funding and the region's degree centrality as an example: On the one hand, funding is expected to positively affect the cooperative performance of regions. However, in similar veins, a high number of cooperative linkages is also likely to increase the probability to raise further funds in the next period (or even immediately). The latter feedback mechanism is the source for the reversed causality problem. In order to minimize this problem, we only take funding volumes allocated throughout the time period 1997-2002, to measure the impact on the region's network position in 2005. A feedback effect is thus rather unlikely to occur (though it is still possible of course if time constant fixed effects are present). The period 1997-2002 was chosen since it equals the funding period in the BioRegio contest (BRC) and thus may be used to indicate the effect stemming from BRC funding on the region's network position. However, one has to note that – although time exogeneity of R&D funding with respect to degree centrality is defined tautologically – this only holds if the latter variable is sufficiently time-varying.[10] A detailed description of the chosen time period for each individual regressor is given in Table 3.

Since data on biotech cooperations is only available for the year 2005, we have to estimate the model in a cross-sectional setting. Surely, we are aware of the problem that an ideal regression design would make use of panel data estimators allowing to control for individual heterogeneity in the data driven by unobserved region specific factors and its potential correlation with the set of regressors ($\mathbf{x}$ and $\mathbf{z}$). Indeed, a visual inspection of selected variables in Figure 4 shows a high degree of regional heterogeneity and spatial clustering. Since the introduction of a vector of regional fixed effects is not feasible in our cross-sectional model, we use spatial filters as a surrogate for regional fixed effects (see Patuelli et al., 2009). In fact, controlling for spatial autocorrelation (SAC) does not only allow to capture omitted variables, but may also correct for self-correlation and/or spatial spillover effects, which can lead to inconsistent or inefficient estimation results (see Cliff and Ord, 1981; Anselin, 1988).

[9] An alternative strategy would be to use instrumental variables with strictly exogenous instruments. However, it is very hard – if not impossible – to find such instruments. Moreover, the IV approach has weaknesses of its own, such as the weak instrumentation problem.

[10] Given that we have only cross-sectional observations at hand, unfortunately this cannot be tested empirically here.

In line with Grimpe and Patuelli (2010), we include an eigenvector-based spatial filter in the model in order to account for the potentially uneven - and spatially correlated - regional distribution of biotech cooperations in Germany. The advantage of the spatial filtering approach compared to alternative spatial regression techniques is that the spatial filtering approach does not require an assumption of normality or other estimation restrictions, and can be straightforwardly applied to regression equations with any underlying distribution (including generalized linear and count data regression models). Here we use an approach developed by Griffith (2000, 2003), which extracts orthogonal and uncorrelated numerical components (eigenvectors) from a projection matrix of an exogenously specified spatial weights matrix $\mathbf{W}$. We take a rook-type binary contiguity weighting matrix, which either takes a value of one if two NUTS3 districts share a common border or is zero otherwise.

Starting point for the generation of candidate eigenvectors is the transformation of the weighting matrix $\mathbf{W}$ according to

$$(\mathbf{I} - \mathbf{11}'/N)\mathbf{W}(\mathbf{I} - \mathbf{11}'/N), \tag{5}$$

where $\mathbf{I}$ is an $(N \times N)$ identity matrix, and $\mathbf{1}$ is an $(N \times 1)$ vector containing ones. The extraction according to eq.(4) results in a set of N eigenvectors (e_i with $i=,\ldots,N$), which have the properties of maximizing spatial autocorrelation while being orthogonal to the previously extracted eigenvectors. These eigenvectors represent all possible patterns of latent spatial autocorrelation implied by the chosen form of $\mathbf{W}$. To reduce the total number of included eigenvectors in the regression equation, we follow Grimpe and Patuelli (2010) and first select a subset of candidate eigenvectors according to the following threshold $MI(e_i) / \max_i \left[MI(e_i) \right] > 0.25$, where $MI(e_i)$ is the Moran's I (MI) indicator for spatial autocorrelation computed based on a generic eigenvector e_i.[11] We then use a stepwise regression design starting from a full model to exclude statistically insignificant eigenvectors in each regression setup and finally test for the joint significance of the remaining eigenvectors by means of a Wald test.[12]

To apply the spatial filtering approach in the case of the disaggregate dyadic regression specification, we need to transform the standard spatial weighting matrix $\mathbf{W}$ into a network weighting matrix $\mathbf{C}$, which extends the two-dimensional space for $(N \times N)$-

[11] This threshold level corresponds to 95% of variance explained in a regression of a generic variable Y on $\mathbf{W}Y$, where the latter is the spatial lag of Y defined as $\mathbf{W}^*Y$.

[12] As Grimpe and Patuelli (2010) have shown, the stepwise elimination of non-significant eigenvectors may tend to result in an overfitted model. We therefore choose a relatively high significance level of 99% for the underlying likelihood ratio test of variable exclusion. Additionally, we only include the set of eigenvectors in the negative binominal part of the ZINB model.

regional tuples with $(i, j \mid i \neq j; i, j = 1, \ldots, N)$ of $\mathbf{W}$ to a four dimensional space with $(n^2 \times n^2)$ possible linkages for $(i, j, r, s \mid i \neq j; r \neq s; i, j = 1, \ldots N; r, s = 1, \ldots N)$.[13] Based on this information, the network weight matrix $\mathbf{C}$ can be constructed as[14]

$$\mathbf{C} = \mathbf{W} \otimes \mathbf{I} + \mathbf{I} \otimes \mathbf{W} = \mathbf{W} \oplus \mathbf{W}. \tag{6}$$

Candidate eigenvector selection and statistical inference for the joint significance of the eigenvectors are then carried out as described above.

The results for the aggregate model specification are shown in Table 4. In column I and II we first estimate a negative binominal model which includes a core set of variables and successively augments the number of regressors. The latter are classified into two broad categories "policy" and regional innovation system ("RIS"). With respect to the policy variables in column I, we only include a set of binary dummies indicating whether the respective NUTS3 region was a part of a winner cluster in the BioRegio contest, participating in a non-winning cluster in the BioRegio contest or was part of a winner cluster in the follow-up BioProfile contest. The basic idea of including these dummy variables is to check for level differences between regions in the *BioRegio* network and outsiders caused by preferential access to R&D funding as well as signalling and mobilizing effects of the contests. As the results show, for the case of BRC winners we indeed observe a positive coefficient of the dummy variable, indicating that BRC winners have - on average - a higher degree centrality. However, the coefficients for the further dummy variables turn out statistically insignificant.

In column II, we then introduce alternative policy variables, namely the total volume of individual and collaborative R&D funding received by regional biotech actors. As the results show, the BioRegio dummy turns out statistically insignificant if we control for the amount of R&D funding in biotechnology received by each region. The policy related effect on the region's cooperative linkages is thus directly related to R&D funding with no further evidence for an additional non-pecuniary signalling/prestige or mobilizing effects. This result also holds, if we move from a negative binominal to a zero inflated negative binominal model in columns III to V. The statistically insignificant result may stem from two counteracting effects being at work: On the one hand, participating in a biotech contest such as BioRegio or BioProfile has a certain mobilizing effect, which may result in new cooperations. Additionally, being a winner typically improves the image of the region and its actors. However, on the other hand, winning regions may show the

[13] As outlined above, to avoid overfitting the model and the use of double counts, in the regression approach we only employ information from the lower triangular part of the matrix of observations.
[14] For a detailed description see, for instance, Chun and Griffith (2010).

tendency towards a closure of their network in terms of dense cliques of strongly interconnected actors. As a result, the net effects of the prominent contests of cooperation in German biotechnology may be small or absent – as observed for our sample data.

Since we are using logarithmic transformations for the set of regressors (except for the binary dummy variables) the obtained regression parameters can be interpreted in a straightforward manner as elasticities.[15] Thus, a 1 percent increase in the volume of collaborative R&D funding leads to a 0.07-0.10 percent change in the degree centrality. Our obtained results support a key finding of Fornahl et al. (2010), 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 so. In our regression exercise, we get similar results for the impact on the region's degree centrality, namely that only collaborative R&D fundings turn out statistically significant in the case of the ZINB.

Looking at the impact of factors from the regional innovation system, the ZINB regression results for the preferred empirical specifications in column IV and V (for the degree and interregional degree centrality, respectively) show that cooperative behavior is positively influenced by the regional knowledge stock, measured in terms of regional patent applications in the field of biotechnology.[16] Also, the percentage share of business start-ups in High-Tech sectors according to the OECD (2010) classification is positively correlated with the number of cooperative linkages. If we look at the number of total linkages, we also observe a positive correlation between the number of biotech firms per region and its total number of links. However, if we only look at interregional linkages (total linkages net of loops), the relationship turns out to be negative, indicating that regions with a critical mass of biotech firms rather engage in internal rather than interregional cooperations. However, the number of intraregional linkages (loops) is positively correlated with the interregional degree centrality, indicating that the former serve as a transmission channel to engage in interregional cooperations. While we observe a positive effect of urbanization forces, measured in terms of population density, on the regional number of biotech linkages (both total and interregional), localization forces proxied by industry specialization and sectoral concentration generally show a negative correlation with the number of regional linkages. This finding is in line with the argumentation in Cantner and Graf (2003), who argue that for

[15] A logarithmic transformation was chosen to control for heteroscedasticity in the sample given that the regional heterogeneity is very large for most variables.
[16] For a definition of biotechnology related IPC classes, see the appendix.

hightech regions the number of cooperations is expected to be the highest for some intermediate degree of specialization.

In the binary probit part of the ZINB, both positive values for individual and cooperative R&D funding reduce the probability of having no cooperative linkages. Thus, the access to public R&D grants may be seen as important prerequisite to engage in R&D cooperations. Also, for regions that do not have any patent application, the probability of having access to the biotech network and cooperative activity is significantly reduced. As the postestimation tests show, the included spatial filter turns out to be statistically significant in all regression specifications. In all cases the LR test rejects the validity of the Poisson regression model in favor of the negative binominal specification. The Vuong test additionally shows that the zero inflated negative binominal model is the best empirical choice for our data settings.

Turning to the estimations setup for the dyadic model in Table 5, we are able to refine the empirical specification with respect to two dimensions: Firstly, we include a new set of dummy variables measuring non-pecuniary effects of the BioRegio and BioProfile contests, which identifies for the following combinations of regions: winner-winner regions, winner-participant regions, winner-nonparticipant regions as well as participant-nonparticipant regions. Secondly, we can control for the geographical distance among regional 2-tuples as a general measure for proximity among the biotech actors.

The obtained empirical results in Table 5 mainly support the findings from the aggregate specification: The different dummy variables do not turn out statistically significant and the policy related funding effect is entirely captured by the average volume of collaborative R&D funding for each regional tuple. The average share of high-tech start-ups as well as the share of knowledge intensive service sector start-ups positively influence the number of cooperative linkages among regions. The same accounts for the average population intensity and the regional human capital stock. A strong sectoral specialization in service sectors tends to lower interregional degree centrality, the same effect holds for the geographical distance between regions, which serves as a strong impediment to cooperative behavior and is pretty much in line with our ex-ante theoretical expectations.

The negative role of distance also turns out statistically significant in the probit part of the ZINB model. This means that an increasing distance between two regions lowers the probability that any interregional cooperation will occur. As for the overall specification in Table 4, also the average volume of individual and collaborative R&D funding and the

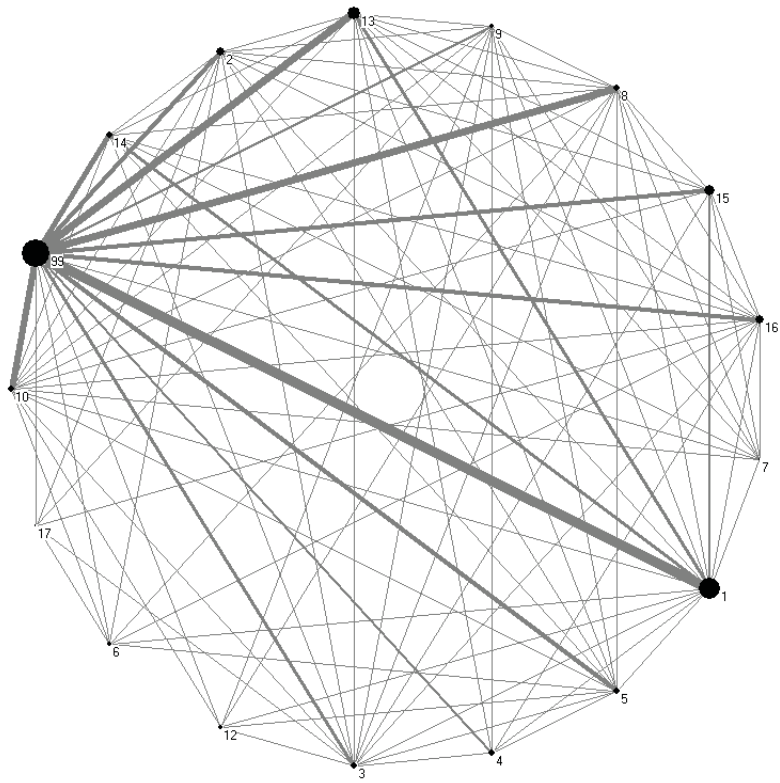
average number of patent applications turn out negative and statistically significant in the probit part of the disaggregate ZINB.

5. Conclusion

In this paper, we have analysed the network formation and its determinants for Germany's biotechnology industry using social network analysis and a regression framework for count data. Our results show that parameters of the underlying regional innovation system as well as policy instruments have an impact on the region's degree centrality as outcome variable in the model. With regard to policy indicators, we find that the volume of collaborative R&D funding is positively correlated with the region's overall and interregional degree centrality, while the amount of individual R&D funding does not seem to matter. Moreover, next to the direct funding effect associated with collaborative R&D subsidies no further non-pecuniary advantages such as prestige, image or mobilizing effects etc. are found. This finding indicates that, on the one hand, cooperative R&D funding is indeed an important policy tool in order to increase the connectivity of actors in an R&D based industry network. On the other hand, we do not get strong empirical evidence that novel cluster based policies such as the BioRegio and BioProfile contests have additional indirect effects on the regional cooperative behavior as typically intended by the design of these policy schemes (e.g., mobilization and signalling effects). Further research thus needs to be conducted in order to carefully evaluate the success of specific contest based policies compared to standard (collaborative) grant schemes.

The inclusion of additional regressors for the role played by the sector-specific and broader regional innovation system, shows that the number of biotech patent applications, the share of regional start-ups in hightech sectors and knowledge intensive service industries as well as population density among other factors are estimated to have a positive effect on the cooperative behavior of regional actors. On the contrary, geographical distance is found to be a strong impediment to engage in interregional cooperative activity. This finding supports the prominent role given to the notion of (geographical) proximity as a key determinant of network formation. Throughout the empirical analysis we have tried to carefully handle potential pitfalls (such as regional heterogeneity and right-hand side endogeneity) when identifying causal effects on regional cooperative behavior. Nevertheless, future research effort is needed to fully understand the complex interrelationship between network features, sectoral as well as locational factors, innovation and economic success.

Figure 1: Cooperative linkages within and between German *BioRegios* in 2005



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

Source: Data from BIOCOM AG (2005), the definition of *BioRegios* is taken from Dohse (2007).

Table 1: Degree centrality and average degree for German *BioRegios*

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

Table 2: Relative importance of internal and external linkages for *Bioregios*

		1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.
		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**

Figure 2: Frequency distribution of the average degree for NUTS3 districts

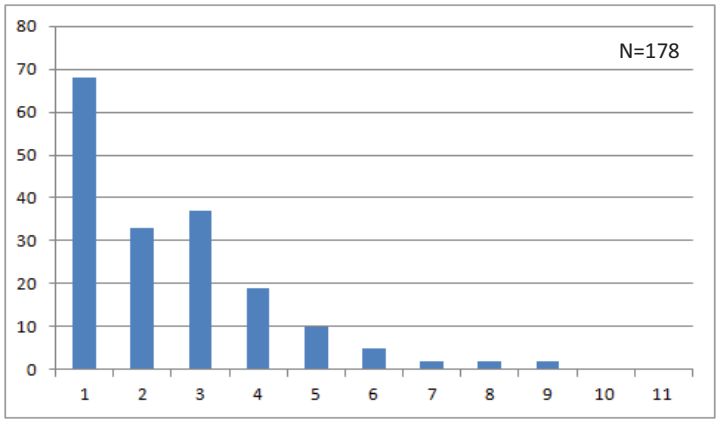
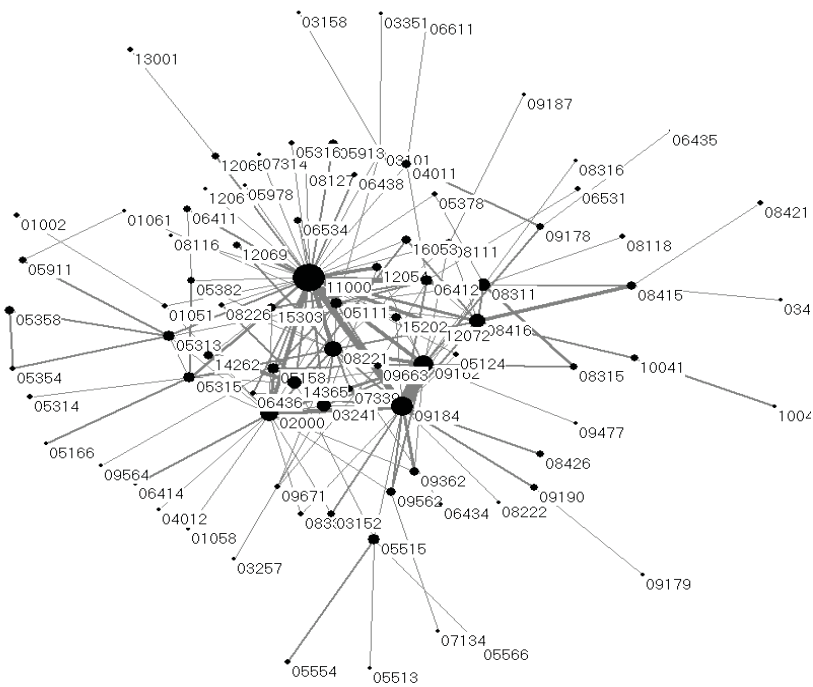


Figure 3: Cooperation linkages within and between German NUTS3 districts



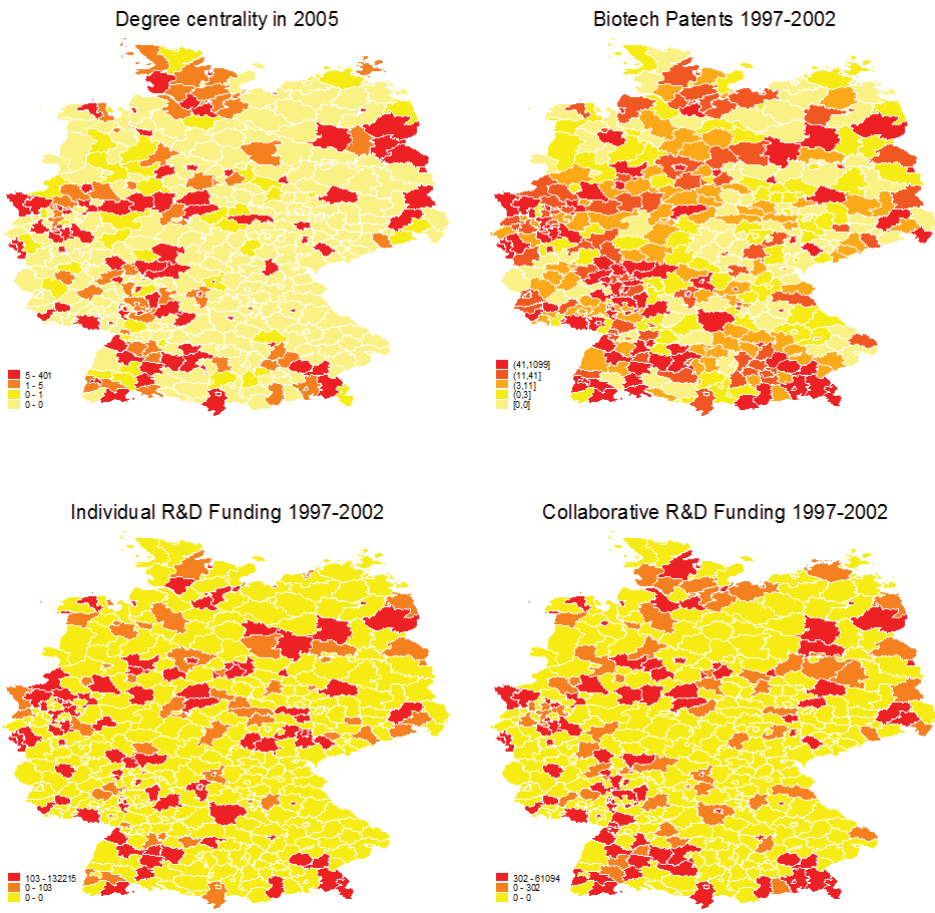
Source: Data from BIOCOM AG (2005).

Table 3: Variable definition and descriptive statistics

Variable	Source	Description	N	Period	Mean	Std. Dev.	Min	Max
Degree	BIOCOM AG	Number of total R&D Collaborations	439	2005	6.332	24.682	0	401
Interreg. Degree	BIOCOM AG	Number of interregional R&D Collaborations	439	2005	5.565	19.859	0	299
No. of Loops	BIOCOM AG	Number of intraregional R&D Collaborations	439	2005	0.768	5.310	0	102
No. of Firms	BIOCOM AG	Number of biotechnology firms	439	2005	2.278	7.127	0	100
Volume of individual R&D Funding	PROFI	Direct funding of biotechnology related R&D individual projects by Federal government (in 1000 €)	439	1997-2002 (sum)	1633.881	9728.613	0	132214.60
Volume of collaborative R&D Funding	PROFI	Direct funding of biotechnology related R&D collaborative projects by Federal government (in 1000 €)	439	1997-2002 (sum)	1082.702	4645.315	0	61094.44
Number of Biotech Patent applications	European Patent Office (EPO)	Weighted number of patent application in biotechnology (for a definition of the Biotech sector based on IPC classes see appendix)	439	1997-2002 (sum)	16.058	46.893	0	576.060
% High-Tech Start-Ups	ZEW Foundation Panel	Number of start-ups in high-tech industries relative to MINT employees (1 = 100 percent)	439	1996-2003 (average)	0.005	0.003	0	0.021
% Medium-Tech Start-Ups	ZEW Foundation Panel	Number of start-ups in medium-tech industries relative to MINT employees (1 = 100 percent)	439	1996-2003 (average)	0.007	0.003	0.001	0.036
% Knowledge intensive Services Start-Ups	ZEW Foundation Panel	Number of start-ups in knowledge intensive services relative to MINT employees (1 = 100 percent)	439	1996-2003 (average)	0.063	0.020	0.021	0.154
% Exports	German Statistical Office	Share of foreign turnover in manu-facturing sector relative to total turnover in the sector (1 = 100 percent)	439	1997-2002 (average)	26.050	13.572	0	96.186

Variable	Source	Description	N	Period	Mean	Std. Dev.	Min	Max
% MINT Employment	Bundesagentur für Arbeit (Federal Employment Agency)	Share of employees trained in mathematics, informatics, natural sciences and technology relative to total employment (in percent)	439	1997-2002 (average)	2.198	1.290	0.450	13.550
Population Density	German Statistical Office	Number of inhabitants per area (in squared kilometers)	439	1997-2002 (average)	514.270	662.510	40.838	3904.829
Sectoral Specialization Manu	Alecke et al. (2006)	Sum of squared deviations in employment shares for NACE3 sectors between regional and national average (manufacturing)	439	1998	713.354	901.530	144.480	8120.970
Sectoral Specialization Serv1	Alecke et al. (2006)	Sum of squared deviations in employment shares for NACE3 sectors between regional and national average (business-related services)	439	1998	254.180	202.685	39.130	2911.670
Sectoral Specialization Serv2	Alecke et al. (2006)	Sum of squared deviations in employment shares for NACE3 sectors between regional and national average (household-related services)	439	1998	129.693	98.588	19.530	609.070
Ellison-Glaeser Index Manu	Alecke et al. (2006)	Employment in sectors with high Ellison-Glaeser-Index (>0.005) relative to total employment in the region (manufacturing)	439	1998	27.741	35.697	0.210	390.832
Ellison-Glaeser Index Serv1	Alecke et al. (2006)	Employment in sectors with high Ellison-Glaeser-Index (>0.005) relative to total employment in the region (business-related services)	439	1998	12.086	36.241	0.054	460.573
Ellison-Glaeser Index Serv2	Alecke et al. (2006)	Employment in sectors with high Ellison-Glaeser-Index (>0.005) relative to total employment in the region (household-related services)	439	1998	6.753	23.263	0.013	279.038
Geographical Distance	BBSR	Driving time (in minutes) between centroids of each NUTS3 district	439	2005	308.744	152.740	0	844.503

Figure 4: Spatial distribution of selected variables in the data



Note: For detailed data descriptions see Table 3. Volume of individual and cooperative R&D Funding as sum for the period 1997 – 2002.

Table 4: Regression results for C^D in the overall network model (linkages for region *i* with all other regions)

Category	Model:	Negativ Binominal		Zero Inflated Negative Binominal		
	Dependent Variable:	Degree	Degree	Degree	Degree	Interreg. Degree
		I	II	III	IV	V
Policy	Volume of individual R&D Funding		0.041***		0.018	0.018
			(0.015)		(0.012)	(0.012)
	Volume of collaborative R&D Funding		0.104***		0.064***	0.06***
			(0.015)		(0.013)	(0.013)
	Dummy BioRegio Winner	1.148**	0.133	0.476	-0.013	-0.10
		(0.524)	(0.373)	(0.293)	(0.231)	(0.227)
	Dummy BioRegio Participant	0.555	0.182	0.662***	0.322*	0.145
		(0.390)	(0.284)	(0.236)	(0.189)	(0.183)
	Dummy BioProfile Winner	0.639	0.198	0.592**	0.051	0.016
		(0.510)	(0.371)	(0.285)	(0.233)	(0.224)
RIS	No. of Firms	0.189***	0.056***	0.06***	0.039***	-0.037***
		(0.037)	(0.018)	(0.010)	(0.007)	(0.014)
	No. of Loops					0.064***
						(0.014)
	Number of Biotech Patent applications		0.134***		0.091***	0.094***
			(0.032)		(0.032)	(0.031)
	% High-Tech Start-Ups	0.311	0.218	0.577***	0.381**	0.41**
		(0.209)	(0.174)	(0.208)	(0.169)	(0.164)
	% Medium-Tech Start-Ups	0.092	0.177	0.024	0.097	0.052
		(0.263)	(0.217)	(0.221)	(0.185)	(0.178)
	% Knowledge intensive Services Start-Ups	0.926**	0.289	0.553*	0.249	0.107
		(0.406)	(0.344)	(0.333)	(0.283)	(0.281)
	% Exports	-0.03	-0.044	-0.051	-0.071*	-0.068*
		(0.050)	(0.044)	(0.042)	(0.037)	(0.036)

Category	Model:	Negativ Binominal		Zero Inflated Negative Binominal		
	Dependent Variable:	Degree I	Degree II	Degree III	Degree IV	Interreg. Degree V
RIS	% MINT Employment	0.365	0.083	0.367*	0.261	0.215
		(0.274)	(0.215)	(0.199)	(0.161)	(0.156)
	Population intensity	0.071	0.167*	0.263***	0.226***	0.242***
		(0.118)	(0.097)	(0.093)	(0.077)	(0.075)
	Sectoral Specialization Manu	-0.004	-0.01	0.069	0.082	0.052
		(0.129)	(0.104)	(0.095)	(0.079)	(0.077)
	Sectoral Specialization Serv1	-0.303*	-0.08	-0.259**	-0.095	-0.047
		(0.164)	(0.138)	(0.125)	(0.109)	(0.107)
	Sectoral Specialization Serv2	-0.218	-0.185	-0.102	-0.127	-0.138*
		(0.139)	(0.114)	(0.101)	(0.085)	(0.082)
	% Ellison-Glaeser Index Manu	-0.397***	-0.325***	-0.195*	-0.237***	-0.23***
		(0.146)	(0.120)	(0.110)	(0.090)	(0.086)
	% Ellison-Glaeser Index Serv1	-0.111	-0.026	-0.342	-0.242	-0.16
		(0.352)	(0.296)	(0.286)	(0.237)	(0.232)
	% Ellison-Glaeser Index Serv2	0.073	0.054	0.271	0.251	0.176
		(0.2854)	(0.2381)	(0.229)	(0.188)	(0.185)
	Constant			6.412***	4.261***	3.547**
				(1.849)	(1.517)	(1.488)
	Probit					
	Volume of individual R&D Funding			-0.092***	-0.079***	-0.083***
				(0.026)	(0.024)	(0.026)
	Volume of collaborative R&D Funding			-0.118***	-0.091***	-0.095***
				(0.024)	(0.020)	(0.022)
	Number of Biotech Patent applications			-0.077***	-0.064**	-0.062*
				(0.027)	(0.031)	(0.032)

Category	Model:	Negativ Binominal		Zero Inflated Negative Binominal		
	Dependent Variable:	Degree I	Degree II	Degree III	Degree IV	Interreg. Degree V
	Constant			-0.839***	-0.70***	-0.744***
				(0.222)	(0.168)	(0.194)
	N	439	439	439	439	439
	Log Likelihood	-730.28	-678.55	-664.17	-637.35	-622.09
	Spatial Filter	43.80**	71.73***	85.52***	130.88***	115.24***
	(P-Value)	(0.04)	(0.00)	(0.00)	(0.00)	(0.00)
	LR (Poisson vs. NB)	831.6***	200.5***	213.1***	62.19***	42.17***
	(P-Value)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	Vuong (NB vs. ZINB)			4.23***	3.61***	3.16***
	(P-Value)			(0.00)	(0.00)	(0.00)

Note: ***,**,* indicate statistical significance at the 1, 5 and 10 percent level. Standard errors in brackets. Specialization and Ellison-Glaeser indices are defined as Manu = manufacturing, Serv1 = business-related services, Serv2 = household-related services. All regressors except the binary dummy variables are specified in logarithmic terms. Spatial Filter tests the joint significance of the included eigenvectors in the regression model. LR (Poisson vs. NB) is a likelihood ratio test for the significance of the dispersion parameter in the negative binominal regression framework (compared to the Poisson). Vuong denotes the Vuong (1989) non-nested test for the Poisson and Zero Inflated Negative Binominal (ZINB) model.

Table 5: Regression results for C^D in the pairwise network model of region tuples (i, j)

Category	Model: / Dependent Variable:	Negativ Binominal Interreg. Degree	Zero Inflated Negative Binominal Interreg. Degree	Interreg. Degree
Policy	Av. Volume of individual R&D Funding			0.004
				(0.026)
	Av. Volume of collaborative R&D Funding			0.199***
				(0.034)
	Dummy BioRegio (Winner x Winner) in same BioRegio	1.162***	0.399	0.331
		(0.425)	(0.286)	(0.296)
	Dummy BioRegio (Winner x Winner) in different BioRegios	0.351	0.205	0.094
		(0.384)	(0.301)	(0.304)
	Dummy BioRegio (Winner x Non-winning Participant) in different BioRegios	0.476**	0.109	0.034
		(0.194)	(0.165)	(0.165)
	Dummy BioRegio (Non-winning Participant x Non-winning Participant) in the same BioRegio	-0.226	-0.531	-0.651
		(0.574)	(0.510)	(0.506)
	Dummy BioRegio (Non-winning Participant x Non-winning Participant) in different BioRegios	0.551**	0.001	-0.070
		(0.260)	(0.240)	(0.239)
RIS	Av. No. of Firms	0.082***	0.053***	0.039***
		(0.006)	(0.007)	(0.007)
	No. of Loops	0.005**	0.002	-0.001
		(0.002)	(0.003)	(0.003)
	Av. Number of Biotech Patent applications			0.042
				(0.084)
	Av. % High-Tech Start-Ups	0.728***	0.612***	0.519***
		(0.166)	(0.200)	(0.195)
	Av. % Medium-Tech Start-Ups	-0.291	0.015	0.057
		(0.208)	(0.252)	(0.246)
	Av. % Knowledge intensive Services Start-Ups	2.428***	1.106***	1.124***
		(0.312)	(0.337)	(0.338)
	Av. % Exports	-0.002	-0.147***	-0.140***
		(0.057)	(0.056)	(0.053)
	Av. % MINT Employment	1.347***	0.432**	0.277
		(0.197)	(0.208)	(0.205)
	Av. Population intensity	0.584***	0.402***	0.345***
		(0.086)	(0.089)	(0.086)
	Av. Sectoral Spezialisation Manu	-0.140	-0.055	-0.028
		(0.091)	(0.091)	(0.091)

Category	Model: / Dependent Variable:	Negativ Binominal Interreg. Degree	Zero Inflated Negative Binominal Interreg. Degree	Zero Inflated Negative Binominal Interreg. Degree
RIS	Av. Sectoral Spezialisation Serv1	-1.193***	-0.486***	-0.400***
		(0.141)	(0.143)	(0.143)
	Av. Sectoral Spezialisation Serv2	-0.344***	-0.328***	-0.329***
		(0.107)	(0.103)	(0.101)
	Av. % Ellison-Glaeser Manu	-0.897***	-0.539***	-0.623***
		(0.102)	(0.105)	(0.105)
	Av. % Ellison-Glaeser Serv1	-0.465	-0.443	-0.286
		(0.284)	(0.308)	(0.297)
	Av. % Ellison-Glaeser Serv2	0.408*	0.435*	0.379
		(0.224)	(0.243)	(0.234)
	Geographical Distance	-0.803***	-0.513***	-0.578***
		(0.049)	(0.054)	(0.061)
	Constant	14.08***	9.76***	8.24***
		(1.731)	(1.805)	(1.821)
	Probit			
	Av. Volume of individual R&D		-0.072***	-0.096***
	Funding		(0.011)	(0.025)
	Av. Volume of collaborative R&D		-0.124***	-0.021
	Funding		(0.011)	(0.028)
	Av. Number of Biotech Patent applications		-0.158***	-0.188***
			(0.023)	(0.052)
	Geographical Distance		0.204***	0.193***
			(0.049)	(0.068)
	No. of Loops		-0.004	-0.004
			(0.002)	(0.004)
	Constant		0.891***	0.210
			(0.263)	(0.401)
	N	96577	96577	96577
	Log Likelihood	-2981.20	-2695.73	-2669.55
	Spatial Filter	368.25***	250.89***	268.53***
	(P-Value)	(0.00)	(0.00)	(0.00)
	LR (Poisson vs. NB)	175.47***	21.57***	1.17
	(P-Value)	(0.00)	(0.00)	(0.14)
	Vuong (NB vs. ZINB)		10.79***	2.40***
	(P-Value)		(0.00)	(0.00)

Note: ***,**,* indicate statistical significance at the 1, 5 and 10 percent level. Standard errors in brackets. Specialization and Ellison-Glaeser indices are defined as Manu = manufacturing, Serv1 = business-related services, Serv2 = household-related services. Average values (Av.) are calculated as $\left(y_i + y_j\right)/2$, where $y_{i(j)}$ is the value for a variable y in Region i(j), respectively. All regressors except the binary dummy variables are specified in logarithmic terms. LR (Poisson vs. NB) is a likelihood ratio test for the significance of the dispersion parameter in the negative binominal regression framework (compared to the Poisson). Vuong denotes the Vuong (1989) non-nested test for the Poisson and Zero Inflated Negative Binominal (ZINB) model.

References

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

Barabasi, A.; Albert, R. (1999): Emergence of Scaling in Random Networks, in: *Science*, Vol. 386(15), pp. 509-512.

Barabasi, A. (2003): Linked, Plume: New York.

Bianconi G.; Barabasi A. (2001): Competition and multiscaling in evolving networks, in: *Europhysics Letters*, Vol. 54(4), pp. 436-442.

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

Boschma, R. (2005): Proximity and Innovation: A critical Assessment, in: *Regional Studies*, Vol. 39(1), pp. 61-74.

Cantner, U.; Graf, H. (2003): Cooperation and Specialization in German Technology Regions, Jenaer Schriften zur Wirtschaftswissenschaft Nr. 4/2003, Friedrich-Schiller Universität Jena.

Cassi, L.; Plunket, A. (2010): The determinants of co-inventor tie formation: proximity and network dynamics, Papers in Evolutionary Economic Geography 10.15.

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*, Vol. 101, No. 3, pp. 523-536.

Cliff, A.; Ord, J. (1981): Spatial and temporal analysis: autocorrelation in space and time, in: Wrigley, N.; Bennett, R. (Eds.): Quantitative Geography, Chapter 10, pp. 104-110.

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

Dickman, S. (1996): Germany Joins the Biotech Race, in: *Science*, Vol. 274, p. 1454.

Dohse, D. (2000): Technology policy and the regions – the case of the BioRegio contest, in: *Research Policy*, No. 29, pp. 1111-1133.

Dohse, D. (2005): Clusterorientierte Technologiepolitik in Deutschland: Konzepte und Erfahrungen, in: *Technikfolgeabschätzung – Theorie und Praxis*, Vol. 1/14, pp. 33-41.

Dohse, D. (2007): Cluster-Based Technology Policy – The German Experience, in: *Industry and Innovation*, Vol. 14, No. 1, pp. 69-94.

Echeverri-Carroll, E. L.; Brennan, W. (1999): Are Innovation Networks Bounded by Proximity?, in: Fischer, M. M.; Suarez-Villa, L.; Steiner, M. (Editors), Innovation, Networks and Localities, Springer: Heidelberg.

Edquist, C. (2000): Systems of Innovations – Their Emergence and Characteristics, in: Edquist, C.; McKelvey, M. (Eds.): Systems of Innovations: Growth, Competitiveness and Employment, Volume 1, MPG Books Ltd.: Cornwall.

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

Engel, D.; Mitze, T.; Patuelli, R.; Reinkowski, J. (2012): Can cluster policies trigger R&D activity? Evidence from German biotech contests, in: *European Planning Studies*, forthcoming.

Falck, O.; Heblich, S. (2008): Modern Location Factors in Dynamic Regions, in: *European Planning Studies*, Vol. 16, No. 10, pp. 1385-1403.

Fornahl, D.; Broekel, T.; Boschma, R. (2010): What drives patent performance in German biotech firms? The impact of R&D subsidies, knowledge networks and their location, Papers in Evolutionary Economic Geography No. 10.09, Utrecht University.

Freeman, L. C. (1978/79): Centrality in Social Networks – Conceptual Clarification, in: *Social networks*, No. 1, Issue 3, pp. 215-239.

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

Glückler, J. (2007): Economic Geography and the evolution of networks, in: *Journal of Economic Geography*, Vol. 7, pp. 619-634.

Griffith, D. (2000): A linear regression solution to the spatial autocorrelation problem, in: *Journal of Geographical Systems*, Vol. 2, pp. 141-156.

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

Grimpe, C.; Patuelli, R. (2010): Regional Knowledge Production in Nanomaterials: A Spatial Filtering Approach, in: *Annals of Regional Science*, forthcoming.

Henkel, J.; Maurer, S. (2010): Network Effects in Biology R&D, in: *American Economic Review: Papers & Proceedings*, Vol. 100, pp. 159-164.

Howells, J. (1999): Regional Systems of Innovation?, in: Archibugi, D.; Howells, J.; Michie, J. (Eds.): Innovation Policy in a Global Economy, Cambridge University Press: Cambridge, pp. 67-93.

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

Krugman, P. (1991): Geography and Trade, Leuven University Press: Leuven.

Lundvall, B. (1992): National Systems of Innovation: Towards a Theory of Innovation and Interactive Learning, Pinter: London.

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

Marshall, A. (1895): Principles of Economics, Vol. 1, 3. Edition, Macmillan and Co.: London and New York 1985.

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

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

Patuelli, R.; Schanne, N.; Griffith, D.; Nijkamp, P. (2010): Persistence of Regional Unemployment: Application of a Spatial Filtering Approach to Local Labour Markets in Germany, Working Paper No. 49-09, Rimini Centre for Economic Analysis.

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

Ter Wal, A. (2011): The Dynamics of the Inventor Network in German Biotechnology: Geographical Proximity versus Triadic Closure, Papers in Evolutionary Economic Geography No. 1102, Utrecht University.

Ter Wal, A.; Boschma, R. (2009): Applying social network analysis in economic geography: framing some key analytic issues, in: *Annals of Regional Science*, Vol. 43, p.. 739-756.

Timmeren, van A.; Röling, L. (2007): Urban and Regional Typologies in Relation to Self-Sufficiency Aiming Strategies, paper presented at the International Conference on "Sustainable Urban Areas", 25.-28. June, Rotterdam, downloaded from: www.enhr2007rotterdam.nl (last access date: 24. April 2012).

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

Wasserman, S.; Pattison, P. (1996): Logit models and logistic regressions for social networks, in: *Psychometrika*, Vol. 61, No. 3, pp. 401-425.

Appendix

Table A.1: List of regions in the BioRegio and BioProfile contests

ID	Name	BioRegio Winner	BioRegio Non-Winner	BioProfile Winner	BioRegio Number
1002	Kiel	0	1	0	10
1003	Lübeck	0	1	0	10
2000	Hamburg	0	1	0	10
13003	Rostock	0	1	0	4
13001	Greifswald	0	1	0	4
3405	Wilhelmshaven	0	1	0	11
3403	Oldenburg	0	1	0	11
4011	Bremen	0	1	0	2
4012	Bremerhaven	0	1	0	2
3241	Region Hannover	0	1	1	2
3201	Hannover	0	1	1	9
3101	Braunschweig	0	1	1	9
3152	Göttingen	0	1	1	9
5124	Wuppertal	1	0	0	13
5111	Düsseldorf	1	0	0	13
5315	Köln	1	0	0	13
5313	Aachen	1	0	0	13
5316	Leverkusen	1	0	0	13
5354	Aachen	1	0	0	13
5358	Düren	1	0	0	13
5314	Bonn	1	0	0	13
6534	Marburg-Biedenkopf	0	1	0	7
6531	Gießen	0	1	0	7
6414	Wiesbaden	0	1	0	14
6412	Frankfurt	0	1	0	14
7315	Mainz	0	1	0	14
6411	Darmstadt	0	1	0	14
6413	Offenbach	0	1	0	14
6436	Main-Taunus	0	1	0	14
6438	Offenbach	0	1	0	14
7314	Ludwigshafen	1	0	0	15
7316	Neustadt a. d. W.	1	0	0	15
8111	Stuttgart	0	1	1	16
8116	Esslingen	0	1	1	16
8221	Heidelberg	1	0	0	15
8222	Mannheim	1	0	0	15
8416	Tübingen	0	1	1	16
8415	Reutlingen	0	1	1	16
8417	Zollernalbkreis	0	1	1	16
8311	Freiburg	0	1	0	3
8421	Ulm	0	1	0	17
9162	München	1	0	0	8
9188	Starnberg	1	0	0	8
9362	Regensburg	0	1	0	12
16053	Jena	1	0	0	6
15202	Halle	0	1	0	5
14365	Leipzig	0	1	0	5
15261	Merseburg-Querfurt	0	1	0	5
15265	Saalkreis	0	1	0	5
15154	Bitterfeld	0	1	0	5
11000	Berlin	0	1	1	1
12065	Oberhavel	0	1	1	1
12069	Potsdam-Mittelmark	0	1	1	1
12072	Teltow-Fläming	0	1	1	1
12054	Potsdam	0	1	1	1

Table A.2: Definition of the Biotech sector based on IPC classes

Patent class	Title
A01H 1/00	Processes for modifying genotypes
A01H 4/00	Plant reproduction by tissue culture techniques
A61K 38/00	Medicinal preparations containing peptides
A61K 39/00	Medicinal preparations containing antigens or antibodies
A61K 48/00	Medicinal preparations containing genetic material which is inserted into cells of the living body to treat genetic diseases; Gene therapy
C02F 3/34	Biological treatment of water, waste water, or sewage: characterized by the micro-organisms used
C07G 11/00	Compounds of unknown constitution: antibiotics
C07G 13/00	Compounds of unknown constitution: vitamins
C07G 15/00	Compounds of unknown constitution: hormones
C07K 4/00	Peptides having up to 20 amino acids in an undefined or only partially defined sequence; Derivatives thereof
C07K 14/00	Peptides having more than 20 amino acids; Gastrins; Somatostatins; Melanotropins; Derivatives thereof
C07K 16/00	Immunoglobulins, e.g. monoclonal or polyclonal antibodies
C07K 17/00	Carrier-bound or immobilized peptides; Preparation thereof
C07K 19/00	Hybrid peptides
C12M	Apparatus for enzymology or microbiology
C12N	Micro-organisms or enzymes; compositions thereof
C12P	Fermentation or enzyme-using processes to synthesize a desired chemical compound or composition or to separate optical isomers from a racemic mixture
C12Q	Measuring or testing processes involving enzymes or micro-organisms; compositions or test papers therefore; processes of preparing such compositions; condition-responsive control in microbiological or enzymological processes
C12S	Processes using enzymes or micro-organisms to liberate, separate or purify a pre-existing compound or composition processes using enzymes or micro-organisms to treat textiles or to clean solid surfaces of materials
G01N 27/327	Investigating or analysing materials by the use of electric, electro-chemical, or magnetic means: biochemical electrodes
G01N 33/53*	Investigating or analysing materials by specific methods not covered by the preceding groups: immunoassay; biospecific binding assay; materials therefore

Table A.2 (continued): Definition of the Biotech sector based on IPC classes

G01N 33/54*	Investigating or analysing materials by specific methods not covered by the preceding groups: double or second antibody: with steric inhibition or signal modification: with an insoluble carrier for immobilizing immunochemicals: the carrier being organic: synthetic resin: as water suspendable particles: with antigen or antibody attached to the carrier via a bridging agent: Carbohydrates: with antigen or antibody entrapped within the carrier
G01N 33/55*	Investigating or analysing materials by specific methods not covered by the preceding groups: the carrier being inorganic: Glass or silica: Metal or metal coated: the carrier being a biological cell or cell fragment: Red blood cell: Fixed or stabilized red blood cell: using kinetic measurement: using diffusion or migration of antigen or antibody: through a gel
G01N 33/57*	Investigating or analysing materials by specific methods not covered by the preceding groups: for venereal disease: for enzymes or isoenzymes: for cancer: for hepatitis: involving monoclonal antibodies: involving limulus lysate
G01N 33/68	Investigating or analysing materials by specific methods not covered by the preceding groups: involving proteins, peptides or amino acids
G01N 33/74	Investigating or analysing materials by specific methods not covered by the preceding groups: involving hormones
G01N 33/76	Investigating or analysing materials by specific methods not covered by the preceding groups: human chorionic gonadotropin
G01N 33/78	Investigating or analysing materials by specific methods not covered by the preceding groups: thyroid gland hormones
G01N 33/88	Investigating or analysing materials by specific methods not covered by the preceding groups: involving prostaglandins
G01N 33/92	Investigating or analysing materials by specific methods not covered by the preceding groups: involving lipids, e.g. cholesterol

Source: OECD (2005), p.32.
Notes: * = Those IPC codes also include subgroups up to one digit (0 or 1 digit). For example, in addition to the code G01N 33/53, the codes G01N 33/531, GO1N 33/532, etc. are included.

Table A.3: Biotech categories in PROFI database

Code: Biotechnology	**Technology field**
K	Biotechnology
I19080	Molecular Bioinformatics

Notes: Own definition according to the technology field classification of the *Leistungsplansystematik des Bundes*. - The following activities have not been considered; "Projektstabskosten" (Code XX XX 90), "Projektbegleiter" (Code XX XX 91), "Beratungsgremien" (Code XX XX 92), "Programmevaluation" (Code XX XX 95).