International Research Networks in Pharmaceuticals: Structure and Dynamics *

Uwe Cantner[†] and Bastian Rake[‡]

April, 2011

Abstract

Knowledge production and scientific research become increasingly more collaborative and international, particularly in pharmaceuticals. We analyze international research networks on the country level in different disease groups. Our empirical analysis is based on an unique dataset of scientific publications related to pharmaceutical research. Using social network analysis we find that both, the number of countries and their connectivity, increases in almost all disease areas. The core of the networks consists of high income OECD countries and remains rather stable over time. We use network regression techniques in order to analyze the dynamics of the networks.

Keywords: International Cooperation, Pharmaceuticals, Research Networks
JEL Classification: R10, O31

*We thank the German Science Foundation (DFG) for financial support. The usual caveats apply.

[†]Friedrich-Schiller-University Jena, Faculty of Economics and Business Administration, Carl-Zeiss-Strasse 3, D-07743 Jena, Germany, and University of Southern Denmark, Department of Marketing and Management, Campusveij 55, DK-5230 Odense M, Denmark, phone: +49-(0)3641-943200, e-mail: uwe.cantner@uni-jena.de

[‡]Friedrich-Schiller-University Jena, Graduate College "The Economics of Innovative Change" (DFG-GK-1411), Carl-Zeiss-Strasse 3, D-07743 Jena, Germany, phone: +49-(0)3641-943-274, e-mail: bastian.rake@uni-jena.de, corresponding author

1 Introduction

Collaboration between different authors and institutions becomes an increasingly more important mode of knowledge generation in almost all scientific disciplines (Wuchty et al., 2007). Particularly in industries with rapidly developing and widely distributed knowledge bases no single actor has the ability to keep pace with the scientific and technological progress in all areas. Consequently, collaboration has become more common. Collaboration networks have been found to be a means by which actors can pool, exchange and develop ideas, knowledge and other resources (Powell and Grodal, 2005, Powell et al., 1996, Powell and Brantley, 1992).

Particularly in the pharmaceutical industry innovations can be seen as the results of interaction and collaboration between a broad set of different types of agents endowed with complementary knowledge, competencies and other resources (e.g. Pisano, 1991, Orsenigo, 1989). In an environment of a complex, expanding and dispersed knowledge based the locus of innovation, and thus the appropriate level of analysis, is no longer the individual actor but rather the entire network (Powell et al., 1996). The structure of the network and the agents' position within it determine the agents' access to the relevant sources of knowledge and therefore their innovative activities and performance (Kogut et al., 1992).

Based on the literature that shows an increasing importance of network structures and the increasing amount of international research collaboration in pharmaceuticals we explore differences in collaboration patterns on the country level in different areas of pharmaceutical research and their developments over time. We use social network analysis to visualize collaboration networks and to calculate network statistics for different disease groups. Moreover, we analyze endogenous network dynamics, i.e. mechanisms within the network that are responsible for new connections being build up or existing ones being cut off. More precisely, we analyze whether homophily, i.e. similarity of countries, preferential attachment, i.e. the connectedness of countries, or multi-connectivity are the driving factors of tie formation within the networks. In order to investigate the network dynamics we employ multiple regression analysis for dyadic data (Butts and Carley, 2001, Krackhardt, 1988). These techniques have to be used since network data violates the assumption of independence between the observations and thus requires different test statistics. The

quadratic assignment procedure (QAP) provides an appropriate method to evaluate the significance of the coefficients (Hubert, 1987, Krackhardt, 1987).

Our empirical analysis is performed on a unique dataset of publications in scientific journals related to pharmaceutical research. We analyze three periods, 1998 to 2000, 2002 to 2004 and 2006 to 2008. Visual inspection reveals that high income OECD countries are located in the center of the network in all periods and disease areas. Although often connected to the core, only few non-OECD countries manage to become part of the center of the international research community. Our descriptive network statistics indicate increasing international collaboration in almost all disease groups. Our preliminary regression results reveal a positive association between preferential attachment as well as multi-connectivity, in terms of different countries connecting two actors, and the change in the number of collaborations. Homophily in terms of income groups is negatively related to the change in international collaborations.

The remainder of the paper is structured as follows: Section 2 presents related literature on research networks and its dynamics. In section 3 we present the methods and the data used in this paper. Descriptive network statistics and visualizations of selected networks can be found in section 4. Results of our regression analysis are presented in section 5. Finally, section 6 concludes.

2 Related Literature

2.1 Research Networks

A network in an economic sense is composed of heterogeneous actors, the relationships among them and other contextual features that affect actors' behavior and decisions as well as the generation and application of knowledge. Concerning the actors involved, many network studies focus on the organizational rather than on the personal, regional, or international level. Regardless which level of analysis is chosen, actors differ from each other in many respects. They have different knowledge and competencies, different rules of action and different incentives and motivations. They are linked among each other through a web of different relationships, including formal links, e.g. contractual

cooperation agreements, as well as less formal relationships, such as joint membership in a community of practice or a regional economy, and all kinds of intermediate relations (Powell and Grodal, 2005, McKelvey et al., 2004).

With respect to a more informal mode of relationships among actors, namely scientific collaboration, there is numerous evidence for an increasing number of co-authored research. This trend towards scientific collaboration has been found in a broad set of disciplines and across different periods (Wuchty et al., 2007, Wagner-Döbler, 2001, de Solla Price, 1963). These studies suggest that the interconnectedness of authors and institutions has considerably increased during the last decades. The increase in scientific collaborations is not restricted to the national level. Adams et al. (2005) show on a large sample of publications originating in U.S research universities, that national and international collaborations increased from the 1980s to the late 1990s. These results are in line with many other studies pointing out the increasing amount of international scientific collaboration in Europe (e.g. Mattsson et al., 2008, Frenken, 2002, Okubo and Sjöberg, 2000). Hence, co-publication networks reveal an expansion in the number of involved countries and the connections among each other. However, not all countries are connected to the core of countries and some are grouped in otherwise disconnected clusters. Over time the global scientific network has become less centralized with new regional hubs emerging (Wagner and Leydesdorff, 2005a).

Increasing collaboration has not only been observed in science but also with respect to R&D and innovative activities in general. Hagedoorn (2002) shows an increasing number of R&D alliances since the 1980s. These alliances are geographically concentrated among North America, Europe, Japan, and South Korea. They can be found in a diverse set of industries, such as the computer, semiconductor, the chemical and the footwear industry (e.g. Boschma and ter Wal, 2007, Ahuja, 2000, Saxenian, 1991). Moreover, collaborative R&D activities show an increasing level of internationalization (Guellec and van Pottelsberghe de la Potterie, 2001, Granstrand, 1999).

In the pharmaceutical industry the R&D process is based on a diverse set of knowledge from different scientific disciplines. The rapid growth of the knowledge base and

its dispersion among a broad variety of actors implies a pronounced trend towards collaboration and network formation. Therefore, innovative activities have been organized in a new organizational form as network of collaborative relations among a diverse set of different actors (Powell et al., 2005, McKelvey et al., 2004). The economic literature presents different interpretations of the motivation, nature, structure, and functions of the observed networks. According to Gambardella (1995) and Arora and Gambardella (1994) collaborations are a new form of organization in response to an increasingly codified and abstract knowledge base. Other interpretations see the industry structure as a transient phenomenon or stress that innovations are the outcome of interaction and collaboration among actors with complementary resources and competencies (e.g. Pisano, 1991, Orsenigo, 1989).

On the organizational level numerous studies described and visualized the growth of R&D partnerships between different types of actors, including established pharmaceutical companies, biotechnology firms, universities, public research institutes, and venture capitalists (e.g. Roijakkers and Hagedoorn, 2006, Powell et al., 2005). Much less emphasis has been put on the international dimension of the collaboration networks. On the country level the network of international R&D projects based on patent data reveals the central role of U.S. based organizations for connecting pharmaceutical research originating in different countries (Owen-Smith et al., 2002).

The international dimension of collaboration in the pharmaceutical industry is particularly pronounced when biotechnological knowledge is involved and regionally clustered actors extent their collaboration beyond national borders (Cooke, 2006). This tendency is reinforced by the fact that biotechnology and pharmaceutical companies locate R&D facilities outside their home countries, connect to a considerable number of international research partners, and source knowledge at a global scale (Tijssen, 2009, Gassmann and von Zedtwitz, 1999). Publication data reflects these observations. In almost one quarter of corporate research publications institutions from at least three world regions are involved (Calero et al., 2007).

2.2 Network Dynamics

Based on the increasing importance of international scientific collaboration, we analyze changes in the collaboration networks over time. The notion of change in evolutionary economics emphasizes processes that lead to a transformation of the economy and its subsystems from within (Witt, 2008, Schumpeter, 1912). Thereby, future events are not independent from past events and the sequence of events influences the outcome. In the context of collaboration networks this evolutionary view implies that the actors' positions and the connections within the network influence future formation and the break-up of ties. Hence the main question in the analysis of network dynamics is how the network structure in previous periods affects interactions among actors, specifically the formation of ties within the network, in subsequent periods (Kenis and Knoke, 2002). There are several theories around which aim to explain the dynamics observed within networks over time. In this paper we concentrate on the concepts of preferential attachment, homophily, and multiconnectivity in order to explain the development of the cross-country collaboration networks in pharmaceutical research.

Real world networks are not randomly generated but show a highly skewed distribution of connections among the involved actors. A small number of actors shows a high number of connections to their counterparts within the network whereas the vast majority of actors has relatively few connections. The distribution of the actor connectivity in real world networks frequently follows at least asymptotically a scale-free power law (Barabási, 2003, Barabási and Albert, 1999). Networks expand through the addition of new actors and already connected actors may build up new connections. The concept of preferential attachment is used to explain the process of growth and intensified collaboration within the network with the characteristics of the network itself. Following the concept of preferential attachment, new and less connected actors establish ties preferably to well connected incumbents. Put differently, the concept states that highly connected actors at one point in time are more likely to attract new connections in the future. Thus, preferential attachment leads over time to a "rich-get-richer" phenomenon in which early entrants increase their connectivity at the expense of newcomers.

Empirical analyses suggest that the mechanism of preferential attachment provides an

explanation for the network structures observed in scientific co-authorship in different disciplines (Wagner and Leydesdorff, 2005b, Jeong et al., 2001, Newman, 2001). Focusing on the firm-level in the pharmaceutical industry after the emergence of biotechnology, Orsenigo et al. (1998) show that the network of collaborative R&D agreements expands but its structural properties remain rather stable. Particularly, the authors find no deformations of the core-periphery structure and a low propensity to collaborate among firms of similar age. These results indicate that preferential attachment may have been the driving force in the evolution of the network (Ter Wal and Boschma, 2009). On the organizational level Gay and Dousset (2005) find evidence for preferential attachment to central actors in the network of antibodies. Different form the theoretical literature preferential attachment seems not to be linked to the age of the actors but rather to the value of their core competencies.

In most real world network the tendency to connect to highly connected actors is not as high as the theoretical models predict. One reason for this observation is that the number of connections that actors can meaningfully maintain is limited. Furthermore, partnering decisions may be influenced by multiple dimensions of proximity. Consequently actors may be attracted by the those with the highest connectivity but prefer to connect to proximate actors (Boschma and Frenken, 2010). Persons and organizations often build up their connections based on similar characteristics in a broad variety of social and economic relations, e.g. marriage, advice and knowledge transfer (for an overview see McPherson et al., 2001, Freeman, 1996). The theoretical concept of homophily, stating that connections are established based on similarity of the actors involved, provides an explanation for the empirical observations. Tie formation based on similarities within the network can be based on restricted opportunities to connect to dissimilar actors induced by the group an actor belongs to and by homophilous preferences (McPherson and Smith-Lovin, 1987). The underlying reasoning of the homophily mechanism is that actors that share similar attributes are more likely to develop characteristic-based trust and to participate in trust-based activities (cf. Zucker, 1986). A high level of similarity among the actors of a network promotes mutual understanding and thus the frequency and intensity of communication and interaction as well as the joint use of knowledge and other resources increases. Hence, interaction within homogeneous networks is subject to a self-reinforcing

process generated within the network (Rogers, 1995). In order to profit form the frequent interaction suggested by homophily mechanism, networks expand by building up new ties to actors having similar characteristics as the members of the network.

In the scientific domain women have been found to collaborate more often with other women and researchers in general tend to connect preferably to people in their own work group (Bozeman and Corley, 2004). Empirical evidence suggests that partnering choices in science are not the only collaborative environments in which homophily may play a role. Ruef et al. (2003) show that the composition of entrepreneurial founding teams is strongly influenced by homophily based on achieved and ascribed characteristics. In contrast to the individual level evidence on the organizational level seems to be less clear. In a study inter-organizational alliances in the German stock photography Glückler (2010) finds that organizational homophily is a relatively weak explanation for the formation of new strategic alliances. Moreover, his results suggest that dissimilarities among the organization may also drive network formation. In the biotechnology industry however, alliance formation is related to homophily (Kim and Higgins, 2007).

Network formation based on preferential attachment and homophily has been contrasted by the multi-connectivity hypothesis. This concept proposes the establishment of multiple connections among the actors of a network through both, direct interaction and intermediaries, driven by a preference for relational diversity (Powell et al., 2005). Networks expand through the establishment of a broad set of independent linkages among the actors. The process may be self-reinforcing since actors that are more diversely linked are more likely to attract more new connections over time than their less diversified counterparts. Hence, a cohesive network structure can evolve.

Empirical evidence shows that the mechanism of multi-connectivity is best suitable to provide an explanation for the formation of strategic alliances in the German stock photography market. The results suggest that two firms are more likely to engage in a partnership if they are connected via third parties (Glückler, 2010). Based on a sample of alliances in life sciences between different types of actors (Powell et al., 2005) find support for the multi-connectivity hypothesis. Their results indicate that the likelihood of new

alliances formation is higher among those actors, that are more diversely connected to each other in the previous period.

Based on the previous literature we find that different mechanism can provide explanations for the observed endogenous network dynamics in real world networks. However, empirical studies show that different mechanisms may be relevant at the same time and that there is no clear-cut explanation which mechanism may explain the network dynamics in the network of international collaborations in pharmaceuticals. Therefore, we aim to analyze the relationship between three alternating mechanisms, preferential attachment, homophily, and multi-connectivity, and the formation and break-up of research collaborations on the country level.

3 Data and Research Methodology

3.1 Social Network Analysis

Social network analysis has been increasingly applied in economics to analyze inventor and co-author networks (Cantner and Graf, 2006, Breschi and Catalini, 2010), knowledge spillovers, and the development of technologies (Mina et al., 2007, Verspagen, 2007). In our study we use social network analysis to illustrate cross-country collaboration patterns in different subfields of pharmaceutical research. The methodology has been mainly developed by anthropologists, sociologists and researchers in social psychology in collaboration with mathematicians, statisticians, and computer scientists. The concept of social networks is based on the assumption of the importance of relationships among interacting units. Beyond this aspect there are four additional paradigmatic properties characterizing social network research. Behavior is seen as interdependent, relational ties are means of resource transfer, the network structure provides opportunities and constraints for individual actions, and the network structure illustrates lasting patterns of relationships (Wasserman and Faust, 1994).

Following these basic characteristics we can define a network as a finite set of actors and their relations among each other. Actors can be defined as discrete individual, corporate, or collective units (Wasserman and Faust, 1994). In the graphical representation

of a network actors are represented as nodes or vertices. Since we aim to analyze cross-country collaborations in the pharmaceutical industry, we refer to countries as the actors in our network. Social ties represent linkages among actors. In order to establish ties among countries we use co-publications between different organizations which may or may not be located in different countries. The collection of ties, i.e. co-publications, defines the relations among the different actors or countries. In the graphical representation of the co-publication network relations among nodes are expressed by undirected arcs.

In order to describe the properties of the cross-country cooperation networks in different therapeutic areas, we compute several descriptive statistics. The number of actors describes the number of countries with at least one publication in the respective field. An important characteristic of a network graph is its connectedness analyzed by computing the number of components. It is connected if there is a path between every pair of nodes. This implies, that all pairs of nodes in the graph can be reached through some path, regardless of its length. Nodes in a disconnected graph can be split up into different subgraphs, the so-called components, which are not connected among each other. A component is a maximal connected subgraph (Wasserman and Faust, 1994). To further examine this property we calculate the size of the largest component and the number of isolated, i.e. disconnected, nodes.

The density of a graph describes the general level of linkages among its nodes. The density is defined as the actual number of connections (lines or edges) of a graph divided by the maximal possible number of lines.

$$\Delta = \frac{\sum d\left(n_i\right)}{g\left(g-1\right)} \tag{1}$$

Where g is the group size, i.e. the number of nodes in the graph, and $d\left(n_i\right)$ is the degree of node i. The degree of a node represents its actual ties to other nodes. The density can take values between 0 and 1. Since it is an average one has to be careful with its interpretation because the variation of the number of ties may be very high. The density of a graph is influenced by the number of isolated nodes since they have by definition a degree of zero.

The mean nodal degree $\bar{d}$ reports the average degree, i.e. the average number of ties

of a node n_i, of all actors in the network.

$$\bar{d} = \frac{\sum_{i=1}^{g} d(n_i)}{g} \tag{2}$$

We can transform the mean degree $\bar{d}$ into the density Δ by dividing it with $g-1$. Actors can be defined as central if they are involved in many relationships within the network. We calculate different centrality measures indicating to which extend actors show high or low levels of centrality and how heterogeneous actors' centrality scores are distributed. One of the simplest definitions of actor centrality states that central actors have to be actively engaged in the network and thus possess a high number of linkages to other actors. Following this idea many researchers used the degree of an node as centrality measure on the individual basis (see Freeman (1979) for an overview):

$$C_D(n_i) = d(n_i) \tag{3}$$

Since this measure depends on the group size g it has to be standardized in order to use it for comparisons across different networks.

$$C'_D(n_i) = \frac{d(n_i)}{g-1} \tag{4}$$

In accordance with the definition of prominence by Knoke and Burt (1983) an actor with a high centrality level is among the most visible ones in the network, being directly connected or adjacent to many others. Actors with low degrees are peripheral to the network and thus less active in the relational process and the information flows. In an extreme case an actor may be complete isolated.

In accordance with Freeman (1979) we can use the measure of actors' degree centrality to construct a general index of graph centralization:

$$C_D = \frac{\sum_{i=1}^{g} [C_D(n^*) - C_D(n_i)]}{\max \sum_{i=1}^{g} [C_D(n^*) - C_D(n_i)]} \tag{5}$$

In the numerator $C_D(n_i)$ refers to the g actor degree indices and $C_D(n^*)$ to the largest observed degree index. Degree centralization of a graph can be expressed by the variation in the actor's degrees divided by the maximum observed degree variation. In the numerator $C_D(n_i)$ refers to the actor degree indices, where $C_D(n^*)$ expresses the largest observed value. The denominator can be expressed directly by $(g-1)(g-2)$ (cf. Freeman, 1979).

$$C_D = \frac{\sum_{i=1}^{g} [C_D(n^*) - C_D(n_i)]}{[(g-1)(g-2)]} \tag{6}$$

Equation 6 gives an index of how centralized the degree of the network's set of actors is. Moreover, it can be interpreted as a measure of dispersion of the actor's degree indices since the latter ones are compared to the maximum value. The degree centralization index equals its maximum value of one if one single, central, actor is related to all other $g-1$ actors, who themselves only interact with the central actor. This is precisely the situation we can find in an ideal star graph. The minimum value of zero is attained if all degrees are equal. This is the case in a regular graph which would correspond to a circle graph (Wasserman and Faust, 1994).

Interactions between non-neighboring nodes are likely to depend on other actors, particularly those lying on the path between the two. The latter ones may play a control or intermediary role concerning the interactions between the other nodes which can be highly valuable for the entire network. The betweenness centrality of a node measures the extend to which this node can be seen as a gatekeeper or broker in the network. This idea has been used to construct the measure of betweenness centrality, which can be considered as the probability that a path within the network takes a particular route. The underlying assumptions are that all lines have equal weight and that the shortest path is used. Freeman (1977) operationalized the idea as the actors' betweenness index which is the sum of all the estimated probabilities over all pairs of actors not including the ith actor:

$$C_B(n_i) = \sum_{j<k} \frac{g_{jk}(n_i)}{g_{jk}} \tag{7}$$

With i being distinct from j and k, let $g_{jk}(n_i)$ denote the number of shortest paths linking actors j and k containing actor i. The probability that two actors are linked by an distinct actor i is given by $g_{jk}(n_i)/g_{jk}$. The index can be standardized so that it takes values between 0 and 1 and can be compared between among different actors and networks.

$$C'_B(n_i) = \frac{2 * C_B(n_i)}{(g-1)(g-2)} \tag{8}$$

Unlike the closeness induces the the betweenness measures can be computed even if the graph is not connected.

The application of group betweenness centralization measures allows us to compare different networks with respect to the variation of the actor's betweenness. According to

Freeman (1979, 1977) we can express the group betweenness centralization index as

$$C_B = \frac{2\sum_{i=1}^{g}\left[C_B\left(n^*\right) - C_B\left(n_i\right)\right]}{\left[(g-1)^2(g-2)\right]} \tag{9}$$

In the numerator $C_B(n_i)$ represents the actor betweenness index and $C_B(n^*)$ its the largest realization. The denominator is the numerator's largest possible value. The index reaches its maximum value of one in a star network, whereas the minimum value of zero is reached if all actors have the same betweenness, i.e. in case of a line graph.

Within a network a path can be characterized as a walk through the net where all lines and all nodes are distinct. The length of a path is its number of lines. The average path length is defined as the average number of lines along the shortest paths between all nodes of the network.

$$L = \frac{1}{g*(g-1)} * \sum_{i \neq j} d_{ij} \tag{10}$$

Where d_{ij} denotes the shortest path between the nodes i and j. The average path length is a structural property of network graphs to determine whether a network fits the small world properties or not (Watts and Strogatz, 1998).

Another indicator that can be used to test the networks' small world properties is the clustering coefficient or transitivity. The intuition behind this measure is the question if two actors that are both connected to a third one interact among each other, too. Accordingly, the clustering coefficient measures the degree to which the nodes of the network tend to cluster together which can be interpreted as the cohesion of the network. A triad involving the actors i, j and k is transitive if i is connected to j as well as j to k and i to k (Wasserman and Faust, 1994). For the entire graph we can compute the global clustering coefficient as the ratio of the number of triangles N_Δ and the number of connected triples N_3 in the graph (Watts and Strogatz, 1998).

$$CC = \frac{3 * N_\Delta}{N_3} \tag{11}$$

The clustering coefficient can be interpreted as the probability that two neighbors of an actor in the network are connected.

3.2 Network Regressions

In order to examine the endogenous mechanisms that drive dynamics of the cross-country collaboration network in pharmaceuticals not only on an descriptive basis we use multiple regression techniques for dyadic data (Butts and Carley, 2001, Krackhardt, 1988). Following Krackhardt (1987), we can describe the relations within a network by a $n \times n$ adjacency matrix, Y in equation 12. The elements of the matrix, $y_{i,j}$, equal zero if there is no relation between actor i and actor j and equal to any other value otherwise. Thus, the values of $y_{i,j}$ indicates the strength of the relation between both actors.

$$Y = \begin{pmatrix} 0 & y_{1,2} & \dots & y_{1,n-1} & y_{1,n} \\ y_{2,1} & 0 & \dots & y_{2,n-1} & y_{2,n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ y_{n,1} & y_{n,2} & \dots & y_{n,n-1} & 0 \end{pmatrix} \tag{12}$$

For the use in regression techniques, the adjacency matrix Y is transformed into a vector form, without the diagonal elements.

$$y = \begin{pmatrix} y_{1,2} \\ y_{1,3} \\ \vdots \\ y_{n,n-1} \end{pmatrix} \tag{13}$$

Applying this transformation to all variables leads to the generalized regression equation for undirected relations (cf. Cantner and Graf, 2006).

$$y_{ij} = \alpha + \beta' x_{ij} + \epsilon_{ij} \text{ for all } i < j \tag{14}$$

In equation 14 the dependent variable y_{ij} may refer to the amount of collaboration between i and j or as in our analytical framework to the change in the number of collaborations. x is a matrix containing the explanatory variables related to the actor pair i and j. This model can be estimated using standard OLS regression techniques. The coefficients are interpreted in the usual way.

Social network data requires different techniques to examine the significance of the coefficients since the problem of structural autocorrelation frequently appears either in rows or columns of the network matrix (Krackhardt, 1987). Thus, significance levels conventional test statistics may provide misleading standard errors and significance levels.

The quadratic-assignment-procedure (QAP) has been found to be an appropriate method to derive more correct inferences concerning the significance of the model's coefficients (Hubert, 1987). QAP provides a general, permutation-based, nonparametric test of the significant relation of two structures (see among others Hubert and Schultz, 1976, Mantel, 1967). The general idea of QAP is to generate the reference distribution against which the coefficients are compared by random permutation of original data matrix' rows and columns. All rows and columns of the matrix are identically permuted which ensures that the structure of the matrix remains unchanged except for those referring to the order of the objects within the matrix (Dekker et al., 2007, Nagpaul, 2003).

The QAP procedure has been found to be quite robust against autocorrelation encountered in network data. Since individual parameter estimates may be biased under multicollinearity, we use a multicollinearity robust version of the QAP procedure proposed by (Dekker et al., 2003). Using QAP regression models, the original regression model is then re-estimated a large number of times. In our study we use 10000 replications of this procedure since this number allows for a sufficient approximation of the reference distribution (cf. Jackson and Somers, 1989).

3.3 Data

Our empirical analysis is performed on an unique dataset of publications in scientific journals related to pharmaceutical research. It was constructed by using different data sources in the following way: First, a list of 251 medical indications was drawn from the BioPharmInsight database.[1] Each indication represents a condition, disease or symptom which allows for the development of a particular procedure or treatment. Each indication is exclusively assigned to one out of 15 therapeutic areas which correspond to a system of an organism or a general diseas group.[2] Therefore, indications assigned to one and the same therapeutic area are considered to be more related than indications that belong to different therapeutic areas.

The list of medical indications was used to conduct a keyword search in the Web of

[1] http://www.infinata5.com/biopharm/

[2] Table 4 provides an overview of the therapeutic areas included in the dataset.

Science databases (WoS). The WoS consist of seven databases containing information gathered from an extensive number of journals, books, book series, reports and conferences. In the case of the Friedrich-Schiller-University of Jena it is hosted by Thomson Reuters. Among these databases the most important one is the Science Citation Index Expanded. It is multidisciplinary and indexes more than 6,500 scientific journals and covers 150 scientific disciplines. The Science Citation Index Expanded covers among others the scientific fields of biochemistry, medicine and pharmacology which are of particular interest for our study. The Web of Science databases include information concerning the scientific publications themselves, like the title, the year of publication, the journal, cited references, a categorization of the research fields a publication can be assigned to and further bibliographic information. In addition to this information the Web of Science reports for most articles the authors' affiliations and their address including the country of origin. However, prior to 2008 it is not possible to match authors with their affiliations.

Publications that contain at least one medical indication from in our keyword list in their title have been included in our dataset. In order to refine the results we only take publications included in categories related to pharmaceutical research into account. More precisely, articles assigned to the subcategories "Biochemistry & Molecular Biology", "Biotechnology and Applied Microbiology", "Chemistry, Applied", "Chemistry, Medicinal", "Medicine, Research & Experimental", "Pharmacology & Pharmacy" and "Toxicology" are included in our dataset.[3] We restrict our sample to journal articles and exclude journal publications that are labeled as meeting abstracts, editorials or reviews as well as other non-journal publications. Conference proceedings have not been considered either since they might be of different quality compared to published papers and may be already included as published articles in the dataset. For the period from 1998 to 2008 we obtain 113057 articles. We further restrict our sample to all articles which contain information concerning the authors' affiliations. In total our sample consists of the 111096 journal articles. In order to analyze the development of international scientific collaboration over time, we distinguish three sub periods, 1998 to 2000, 2002 to 2004, and 2006 to 2008. We do not take the years 2001 and 2005 into account in order to have periods of equal length and to have a clear separation among the sub periods.

[3]The subcategories are described in detail at http://scientific.thomsonreuters.com/mjl/.

We extract information concerning the authors' affiliations and their countries of origin and match it with Worldbank income groups in order to have some information concerning the wealth level of the countries in our sample.

Publication data provides the advantage of getting access to highly detailed information included in scientific articles that are usually available for a long time span. However, there are some drawbacks that have to be taken into account in the analysis of co-publication. The most important are that research does not necessarily lead to publications, co-authorship may only partly capture scientific collaboration, the impact of publications differs considerably and publication habits differ among scientific disciplines. Although researchers using co-publication data face the mentioned shortcomings, this type of data has been found to be an appropriate indicator for scientific collaboration if large datasets, concentrated in one scientific field and aggregated on the country level, are used (see e.g. Katz and Martin (1997), Laudel (2002), Lundberg et al. (2006), and Hoekman et al. (2009) for a discussion).

4 International Research Networks

4.1 Network Descriptives and Visualizations

In this section we employ social network analysis to visualize differences in the international collaboration patterns in pharmaceutical research in various therapeutic areas. We subdivide our analysis in two periods, 1998 to 2000, 2002 to 2004 and 2006 to 2008. Cross-country cooperation networks are illustrated with Pajek (see de Nooy et al., 2005) applying the algorithm proposed by Fruchterman and Reingold (1991). Furthermore, we use the igraph by Garbor Csardi and netmodels package by Domingo Vargas for R statistical software to calculate descriptive network statistics. The spatial position of individual countries within the network represent their relative centrality.

The descriptive network statistics presented in table 1 reveal some general trends in the development of cross-country networks of pharmaceutical research. The number of countries participating in the international research community and the relative size of

the largest component, i.e. the largest group of connected countries, increase in all therapeutic areas from the first to the third period. This corresponds to a decrease in the share of isolated countries, whose firms and research organizations do not cooperate with partners outside their home country. However, their absolute number increases in ten therapeutic areas.

Most networks show an increase in their density from the first to the third period, which indicates that the number of realized linkages grows faster than the number of countries. However, the density remains quite close to its minimum value of 0 in all subnetworks. In most networks the increasing trend is not stable, i.e. that the density decreases in at least one period. The highest share of realized compared to possible linkages, 14.1%, is reached in the area of central nervous system research in the first period. The lowest value with 2.4% is observed in dermatology in the same period. With a few exceptions the mean number of other nations a country is connected to is increasing from the first to the third period. We interpret this as a hint that the cross-country collaboration intensity in pharmaceutical research increases over time.

	Number of Actors	Number of Components	Abs. Size Largest Component	Rel. Size Largest Component	Abs. Number of Isolates	Rel. Number of Isolates	Density	Mean Degree	Degree Centralization	Betweenness Centralization	Average Path Length	Clustering Coefficient
Complete_98_00	136	7	130	0.956	6	0.044	0.107	14.397	0.576	0.221	2.070	0.427
Complete_02_04	141	9	133	0.943	8	0.057	0.119	16.723	0.582	0.195	2.045	0.487
Complete_06_08	154	1	154	1.000	0	0.000	0.136	20.779	0.597	0.167	2.068	0.499
NetTA1_98_00	73	11	63	0.863	10	0.137	0.109	7.863	0.530	0.244	2.091	0.449
NetTA1_02_04	84	9	76	0.905	8	0.095	0.120	9.929	0.556	0.221	2.098	0.483
NetTA1_06_08	101	7	95	0.941	6	0.059	0.127	12.673	0.554	0.212	2.092	0.492
NetTA2_98_00	73	15	58	0.795	13	0.178	0.091	6.548	0.449	0.166	2.184	0.443
NetTA2_02_04	84	15	70	0.833	14	0.167	0.082	6.786	0.422	0.157	2.309	0.443
NetTA2_06_08	89	15	75	0.843	14	0.157	0.118	10.382	0.402	0.122	2.226	0.535
NetTA3_98_00	56	7	50	0.893	6	0.107	0.141	7.750	0.495	0.259	2.024	0.512
NetTA3_02_04	68	9	60	0.882	8	0.118	0.123	8.235	0.596	0.286	2.023	0.453
NetTA3_06_08	79	10	70	0.886	9	0.114	0.127	9.899	0.527	0.207	2.082	0.500
NetTA4_98_00	31	20	5	0.161	16	0.516	0.024	0.710	0.117	0.013	1.565	0.000
NetTA4_02_04	32	17	16	0.500	16	0.500	0.054	1.688	0.183	0.113	2.358	0.510

Continued on next page

	Number of Actors	Number of Components	Abs. Size Largest Component	Rel. Size Largest Component	Abs. Number of Isolates	Rel. Number of Isolates	Density	Mean Degree	Degree Centralization	Betweenness Centralization	Average Path Length	Clustering Coefficient
NetTA4_06_08	35	18	18	0.514	17	0.486	0.049	1.657	0.260	0.153	2.418	0.351
NetTA6_98_00	48	16	33	0.688	15	0.313	0.058	2.708	0.406	0.242	2.388	0.297
NetTA6_02_04	54	12	43	0.796	11	0.204	0.084	4.444	0.481	0.307	2.174	0.353
NetTA6_06_08	69	16	53	0.768	14	0.203	0.067	4.580	0.491	0.254	2.146	0.337
NetTA7_98_00	67	16	51	0.761	14	0.209	0.071	4.687	0.364	0.200	2.460	0.451
NetTA7_02_04	68	14	55	0.809	13	0.191	0.083	5.559	0.499	0.242	2.221	0.432
NetTA7_06_08	77	14	64	0.831	13	0.169	0.096	7.325	0.482	0.175	2.185	0.430
NetTA8_98_00	42	14	29	0.690	13	0.310	0.057	2.333	0.401	0.326	2.495	0.282
NetTA8_02_04	44	14	30	0.682	12	0.273	0.056	2.409	0.429	0.319	2.326	0.274
NetTA8_06_08	55	12	44	0.800	11	0.200	0.071	3.855	0.464	0.337	2.314	0.347
NetTA9_98_00	59	14	44	0.746	12	0.203	0.061	3.525	0.276	0.190	2.526	0.383
NetTA9_02_04	55	14	41	0.745	12	0.218	0.065	3.491	0.433	0.265	2.352	0.305
NetTA9_06_08	63	14	50	0.794	13	0.206	0.084	5.206	0.513	0.345	2.287	0.528
NetTA10_98_00	24	11	14	0.583	10	0.417	0.098	2.250	0.415	0.168	1.824	0.425
NetTA10_02_04	28	11	18	0.643	10	0.357	0.074	2.000	0.439	0.271	2.078	0.250
NetTA10_06_08	38	14	25	0.658	13	0.342	0.077	2.842	0.318	0.165	2.307	0.414
NetTA11_98_00	59	15	44	0.746	13	0.220	0.063	3.627	0.399	0.261	2.317	0.314
NetTA11_02_04	64	12	53	0.828	11	0.172	0.082	5.156	0.473	0.315	2.294	0.384
NetTA11_06_08	72	12	61	0.847	11	0.153	0.129	9.194	0.446	0.154	2.086	0.515
NetTA12_98_00	58	8	50	0.862	6	0.103	0.084	4.793	0.603	0.419	2.151	0.282
NetTA12_02_04	56	11	45	0.804	9	0.161	0.110	6.071	0.489	0.262	2.053	0.475
NetTA12_06_08	72	12	61	0.847	11	0.153	0.103	7.306	0.474	0.197	2.168	0.484
NetTA13_98_00	116	13	104	0.897	12	0.103	0.080	9.224	0.458	0.186	2.189	0.359
NetTA13_02_04	121	7	115	0.950	6	0.050	0.109	13.091	0.508	0.206	2.154	0.467
NetTA13_06_08	132	4	129	0.977	3	0.023	0.111	14.576	0.585	0.188	2.104	0.399
NetTA15_98_00	50	15	34	0.680	12	0.240	0.080	3.920	0.384	0.152	2.062	0.352
NetTA15_02_04	52	11	42	0.808	10	0.192	0.102	5.192	0.465	0.194	2.156	0.399
NetTA15_06_08	65	16	50	0.769	15	0.231	0.072	4.585	0.474	0.253	2.274	0.385
NetTA16_98_00	45	12	27	0.600	8	0.178	0.084	3.689	0.293	0.109	2.000	0.433
NetTA16_02_04	44	13	31	0.705	11	0.250	0.122	5.227	0.360	0.159	2.026	0.505
NetTA16_06_08	54	15	39	0.722	13	0.241	0.091	4.815	0.376	0.181	2.082	0.423
NetTA17_98_00	67	7	61	0.910	6	0.090	0.112	7.373	0.447	0.308	2.268	0.465
NetTA17_02_04	62	11	52	0.839	10	0.161	0.106	6.484	0.483	0.203	2.127	0.436
NetTA17_06_08	77	10	68	0.883	9	0.117	0.095	7.195	0.484	0.211	2.277	0.403

Table 1: Network Descriptive Statistics

The degree centralization measure varies between 0.4 and 0.5 in most networks, which indicates that the number of linkages is quite dispersed among countries in the majority of networks analyzed. This finding indicates that some countries collaborate more than others. All betweenness centralization measures are below 0.42 which indicates some dispersion of this measures among the actors in all subnetworks. Table 1 shows that the average path length between countries is rather stable above 2 in most therapeutic areas. In 10 therapeutic areas the clustering coefficient as a measure for coherence of the network increases from the first to the third period which can be seen as another indicator of increasing international collaboration.

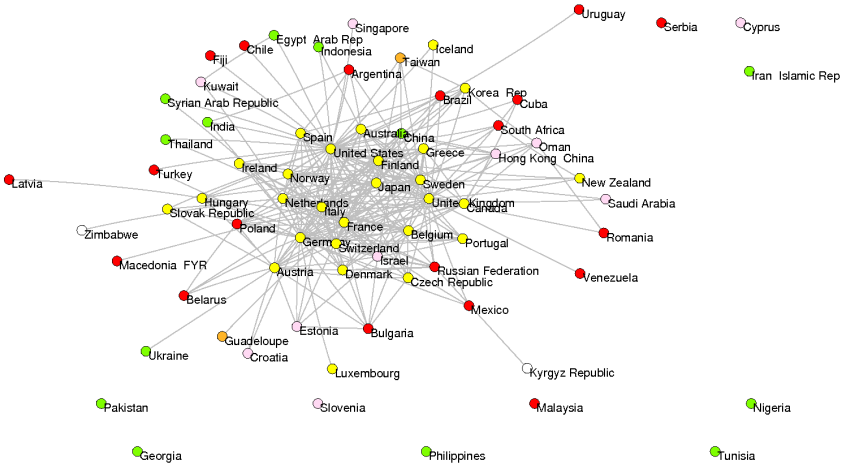
For illustration of the differences among therapeutic areas we choose the international collaboration networks in cancer, infectious diseases and dermatology research. In the case of cancer the network shows a relatively high density and connectedness. For our first period of analysis from 1998 to 2000 cancer publications originated in 73 countries, this number increases to 84 and 101 for the years 2002 to 2004 and 2006 to 2008. The size of the largest component increases from 86.93% of the countries in the first to 94.1% in the third period. This increase is accompanied with a decrease in the absolute and relative number of isolated countries. The density of the network increases over time from 0.109 in the first to 0.127 in the third period indicating an increasing interconnectedness of the countries in the network. Over time, each country in the network is on average connected to more countries. The mean degree rises from 7.863 in the first to 12.673 in the third period. Nevertheless, the degree centralization measure is above 0.5 in all subperiods indicating a quite dispersed distribution of ties among the actors. The decrease of the betweenness centralization measure decreases form 0.244 to 0.212 reveals a decreasing variation of the actors betweenness indices. The average path length stays relatively constant around 2.09 whereas the increase of the clustering coefficient form 0.449 to 0.492 indicates that the network becomes more coherent over time.

Figure 1 illustrates the increasing connectedness of countries in the cancer research network. By visual inspection we see that the most central actors in all three periods can be found among high income OECD member states. Among them are countries that have a rather strong pharmaceutical industry. Particularly these countries are located in the

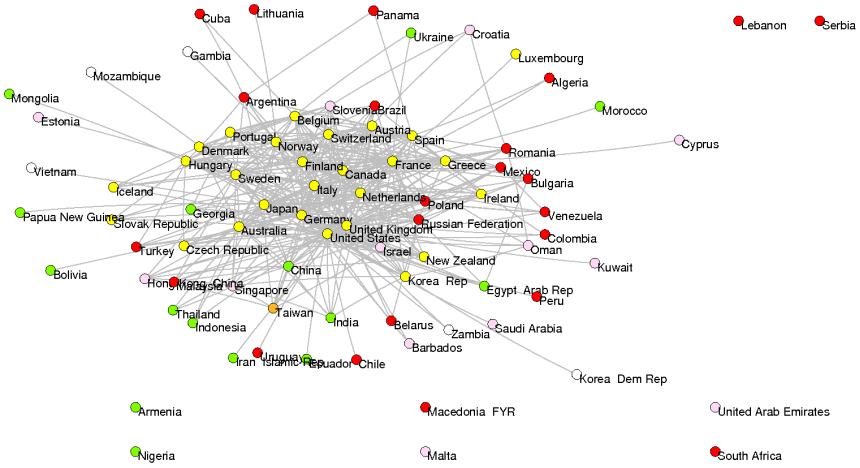
center of the network. Most upper middle income and non-OECD high income countries are located around the core but are connected to it. In the third period we see that China managed to become a central actor in the cancer research network. Several other newly industrializing countries are close to the center of the network, e.g. Brazil, India and Oman. However, most lower middle income and low income countries remained in peripheral positions.

Similar to the cancer network the international research network in infectious diseases shows a relatively high level of participation and connectedness. The number of actors rises from 104 in the first and 121 in the second to 132 in the third period. This development is accompanied with an increase in the relative size of the largest component from 89.7% to 97.7%. Hence the absolute and relative number of isolated countries decreases over time. The density of the network increases from 0.08 to 0.111 indicating that more possible linkages among the countries are realized. The average number of connections a country has build up rises from 9.224 connections in the first period to 14.576 in the third period. The dispersion of actors' degree indices, i.e. the number of connections a country has, increases over time whereas the dispersion of countries' betweenness indices stays rather constant. The average path length decreases slightly. Network cohesion, as indicated by the clustering coefficient, increases slightly from the first to the second period. However, the cohesion is highest in the second period.

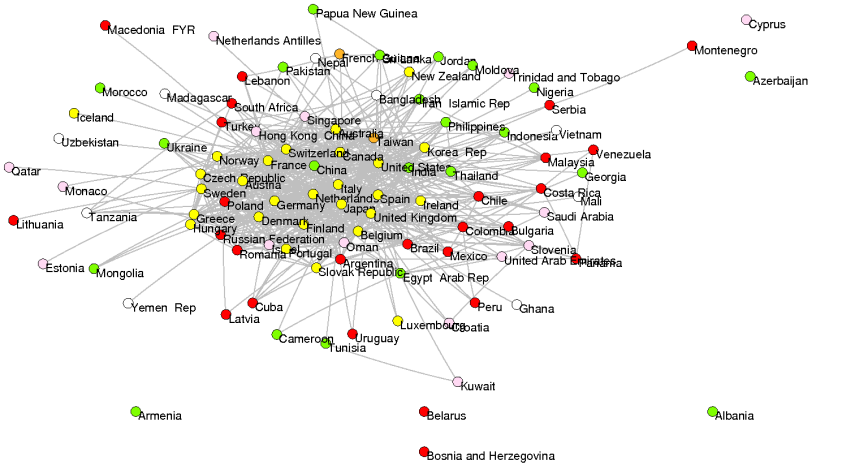
Visual inspection of the infectious diseases networks in figure 2 relatively similar pattern than the cancer network. The core of the network is dominated in all three periods by high income OECD countries. Lower and upper middle income countries are mostly located around the core but are connected to it. However, many of these countries seem to be connected through multiple paths to the core of the network. In the first period Brazil and Thailand have prominent positions within center of the network but they become more peripheral actors in the subsequent periods. In the second period we observe a cluster of Eastern European and former Soviet Union member states that is connected to other participants of the network but indicates intense collaboration among these countries. In the third period however, this cohesive group cannot be identified any longer. In contrast to the cancer network more lower middle income countries are involved in the



(a) 1998-2000



(b) 2002-2004



(c) 2006-2008

Figure 1: Cross-Country Research Networks in Cancer

Income Groups: high income non-OECD (pink), high income OECD (yellow), low income (white), lower middle income (green), upper middle income (red), not classified (orange)

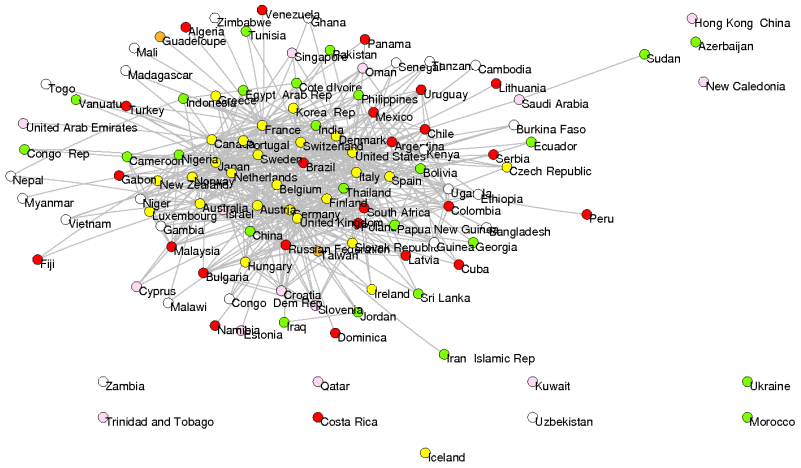
international research network from the first period on, which is most likely associated with the prevalence of infectious diseases in these countries.

In the visualization of international collaborations in dermatology in figure 3 we see that the number of countries engaged in this therapeutic area is considerably lower compared to the cancer and infectious diseases networks. The number of actors in the network rises from 31 in the first to 35 in the third period but collaboration among the countries in the graph seems to be not that intense. We find a consistently large number of different components, most of them consisting of isolated countries. Around 50% of all countries are not connected to any other nation in the network. Hence we find relatively low values for the density and the mean degree, although connectedness rises over time. The degree centralization rises over time as the network becomes more connected indicating that some actors build up more ties than others. The same applies to the betweenness centralization. The average path length and the clustering coefficient increase over time. However, the network remains relatively unconnected in all three periods.

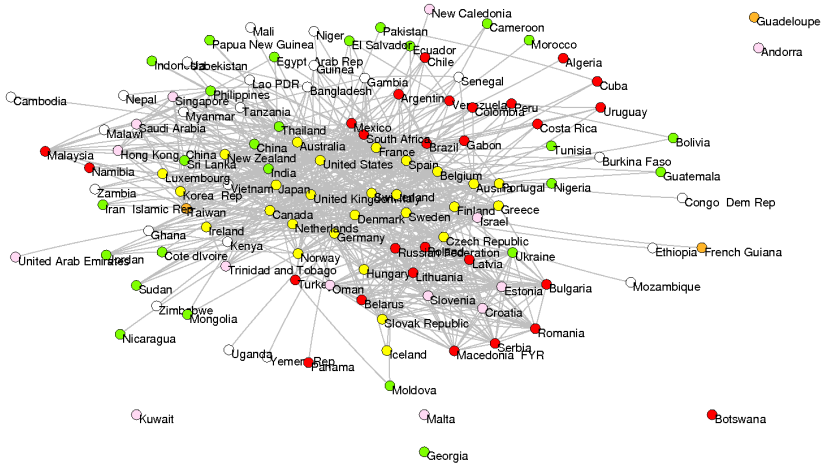
Most of the countries active in the field of dermatology are again high income OECD countries. These countries account for the vast majority of connected actors in the three periods of observations. There are few upper and lower middler income countries that are connected to other nations in one of the three periods. Moreover, we do not find published research originating in low income countries in this field.

4.2 Entry and Exit

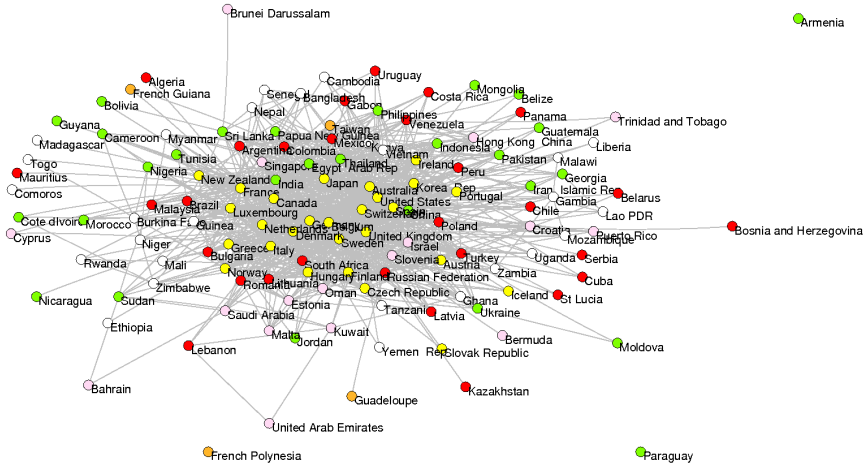
In the previous section network statistics and visualizations indicate intensified collaboration across countries in almost all therapeutic areas. We find that an increasing number of countries is engaged in collaborative pharmaceutical research across borders. However, the network visualizations already indicate that not all countries are persistently engaged in international research projects. In this section we analyze the number of entries, exits and persistently contributing countries in more detail. In doing so, we calculate the mean degree, i.e. the average number of connections an actors has, for the three subgroups mentioned. The connectivity of actors within the network may be associated with their research performance and their decision to leave the network. Based on evidence on the



(a) 1998-2000



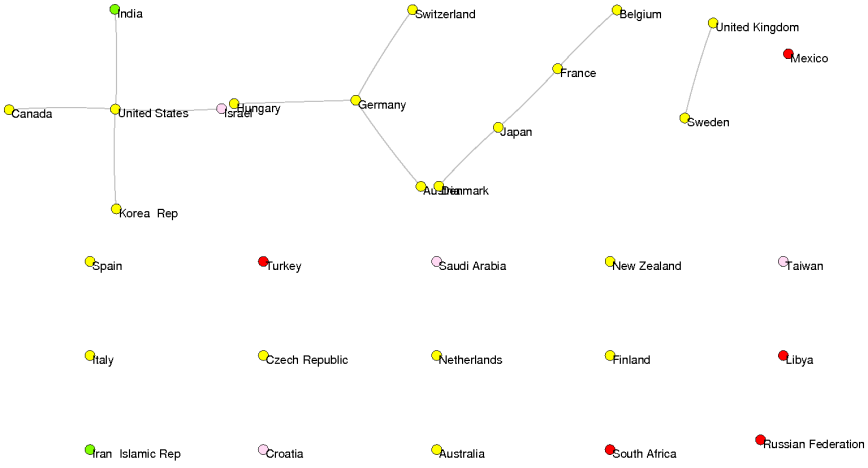
(b) 2002-2004



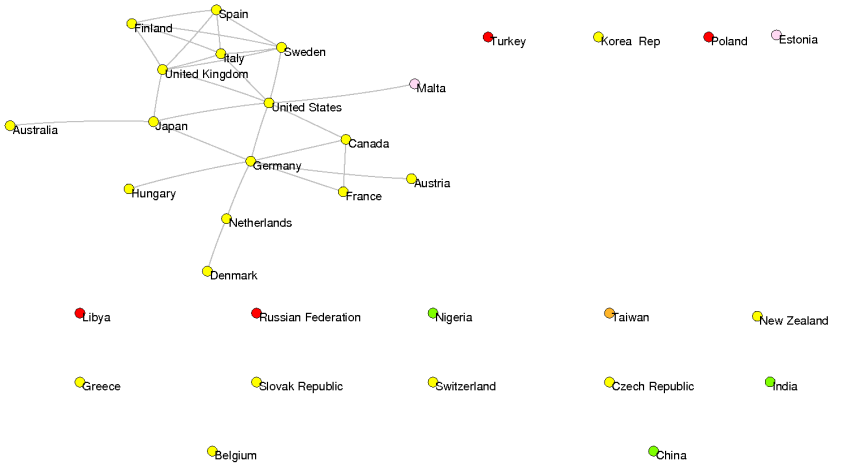
(c) 2006-2008

Figure 2: Cross-Country Research Networks in Infectious Diseaes

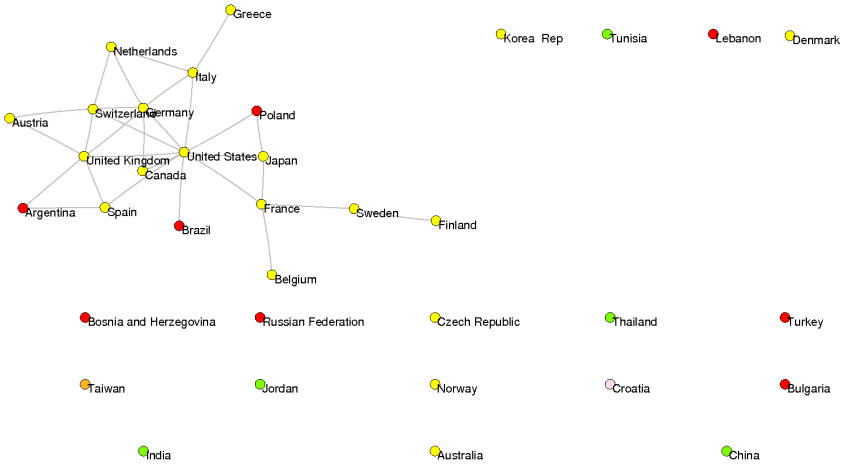
Income Groups: high income non-OECD (pink), high income OECD (yellow), low income (white), lower middle income (green), upper middle income (red), not classified (orange)



(a) 1998-2000



(b) 2002-2004



(c) 2006-2008

Figure 3: Cross-Country Research Networks in Dermatology

Income Groups: high income non-OECD (pink), high income OECD (yellow), low income (white), lower middle income (green), upper middle income (red), not classified (orange)

individual and organizational level we expect countries to leave the network because of a weak position therein, i.e. a relatively low number of connections to other actors (cf. Cantner and Graf, 2006, Powell et al., 1999).

Table 2 reveals a considerable number of entries and exits from the first to the second and from the second to the third period in all therapeutic areas. In thirteen out of 15 therapeutic areas more than ten countries enter and in six therapeutic areas the number of exits is bigger than ten in the first period. The number of entering countries exceeds the number of exits in eleven therapeutic areas. In the third period we find net entry and more than ten entering countries in all therapeutic areas. However, the number of exits increased in six therapeutic areas compared to the previous period.

With respect to the mean degree of each subgroup, entering, exiting and permanent actors, we find considerable differences in all therapeutic areas among these groups. Permanent actors are connected to a by far higher number of other countries than entering and exiting nations. The finding is prevalent for entries and exits from the first to the second and from the second to the third period. We interpret this a a hint that exiting countries leave the international research network because of a relatively weak position of their scientific system in terms of international contacts. Particularly countries that enter in the third period show on average a higher number of connections than the countries exiting. Nevertheless, these group is by far less connected than the permanent actors. The latter increase in 13 out of 15 therapeutic areas their average number of collaborative ties. This finding indicates that these group increasingly engages in international research collaborations.

Our results concerning the connectivity of entering, exiting and permanent contributers fit quite well to our expectations. Countries that leave the network are on average less connected than the permanent actors. Hence, these countries may leave the network due to their relatively weak position. Entrants are particularly in the third period better connected than exiting countries but by far less than permanent actors. The latter ones increase their average number of connections over time. These results indicate that there the international collaboration networks in pharmaceuticals are subject to dynamic tie

	Entries 2002-2004	Mean Degree Entries 2002-2004	Exits 2002-2004	Mean Degree Exits 2002-2004	Permanent Actors 2002-2004	Mean Degree Permanent Actors 2002-2004	Entries 2006-2008	Mean Degree Entries 2006-2008	Exits 2006-2008	Mean Degree Exits 2006-2008	Permanent Actors 2006-2008	Mean Degree Permanent Actors 2006-2008
Complete	20	2.400	15	1.444	121	19.091	22	2.273	9	1.444	132	23.864
TA1	20	1.100	9	1.222	64	12.688	26	3.000	9	1.222	75	16.027
TA2	18	1.056	7	0.909	66	8.348	16	1.375	11	0.909	73	12.356
TA3	20	2.200	8	1.727	48	10.750	22	3.455	11	1.727	57	12.386
TA4	7	0.143	6	0.286	25	2.120	10	0.300	7	0.286	25	2.200
TA6	14	1.357	8	2.222	40	5.525	24	0.958	9	2.222	45	6.511
TA7	15	1.267	14	1.600	53	6.774	19	2.053	10	1.600	58	9.052
TA8	11	0.273	9	0.200	33	3.121	16	1.063	5	0.200	39	5.000
TA9	14	0.714	18	1.091	41	4.439	19	0.947	11	1.091	44	7.045
TA10	10	1.200	6	0.875	18	2.444	18	1.111	8	0.875	20	4.400
TA11	13	1.692	8	1.444	51	6.039	17	2.765	9	1.444	55	11.182
TA12	10	1.000	12	0.429	46	7.174	23	2.348	7	0.429	49	9.633
TA13	15	4.800	10	3.778	106	14.264	20	1.600	9	3.778	112	16.893
TA15	12	1.583	10	2.667	40	6.275	19	1.158	6	2.667	46	6.000
TA16	8	0.500	9	0.333	36	6.278	13	1.000	3	0.333	41	6.024
TA17	14	0.857	19	1.222	48	8.125	24	2.708	9	1.222	53	9.226

Table 2: Entries, Exits and Permanent Actors

formation. However, with the descriptive analysis presented so far we cannot assess if these dynamics can be explained by mechanisms from within the network.

5 Empirical Results Network Regressions

5.1 Variables

We present an overview of the variables and controls used in our network regression models in table 5. The dependent variable is the change in the number of total collaborations between two countries between period 2 and period 3. More precisely, we calculate the number of collaborations for each pair of countries in period 3 and subtract the number of collaborations in period 2. The number of co-publications between each pair of

countries is calculated based on author affiliations. We use full counting which leads to a co-publication count of one for each pair of countries involved in a publication. We include each pair of countries only once in our analysis since co-publications represent indirected links.

With respect to the independent variables, we draw upon multiple measures in order to test the different mechanisms of endogenous network dynamics presented in section 2.2. Following Glückler (2010) we use differences in countries' degree centrality scores lagged by one period as proxy for preferential attachment (*DiffDegreeCentral*). This measure refers to differences in the visibility of countries in the research network. The number of prior ties has been used as another indicator for an accumulative advantage based on preferential attachment (cf. Powell et al., 2005). Therefore, we include the number of previous collaborations in period 1 and 2 (*PrevExp*) in our analysis.

Homophily is reflected by the variable *SmIncomeGr* indicating whether the two collaborating countries belong to the same Worldbank income group, i.e. they have comparable wealth levels. Multiconnectivity is captured by the point connectivity for each country pair lagged by one period (*PointConnectivity*). This measure indicates the number of other countries that have to be removed from the network in order to disconnect two collaborating countries. Moreover, we use the number of shortest paths between two countries in the network with a lag of one period (*GeoCount*) as a further proxy for multiconnectivity.

5.2 Regression Results

In table 3 we present preliminary results of our regression analysis on an aggregated level, i.e. we do not distinguish among the different therapeutic areas. This analysis may deliver some insight which mechanisms drive the formation and the break-up of ties within the network. Network correlations of the independent variables can be found in table 6.

With respect to preferential attachment as a driver of tie formation, we find a positive and significant coefficient for both variables, (*DiffDegreeCentral*) and (*PrevExp*). This indicates a positive relation between differences in the degree centrality of actors in the

previous period and the change in the number of collaborations. We interpret this as a hint that peripheral actors connect to highly connected ones as described by the mechanism of preferential attachment. Previous collaborations among countries are positively related to the number of research collaborations in period 3. These results indicate that preferential attachment based on an accumulative advantage from prior research may be prevalent, too.

Homophily in terms of countries being in the same income group is negatively and significantly related to the formation and break-up of research collaborations. This finding indicates that countries that are in the same income group change their intensity of collaboration to a lesser extent than those in different income groups. We see this as a hint that heterophily rather than homophily in terms of income groups is associated with the formation and break-up of collaborations on the country level.

We analyzed whether multi-connectivity is suitable to explain changes in the number of research collaborations on the country level. We find a positive and significant coefficient for *PointConnectivity*. This finding suggests that changes the intensity of collaborations are positively related to the number of countries that indirectly connect two actors. Put differently, the intensity of collaboration may change due to knowledge flows the partners receive through other collaborations. The coefficient for *GeoCount*, i.e. the number of shortest paths, has a negative sign. However, the results is not significant at conventional levels. Hence, our results indicate that multi-connectivity in terms of the number of shortest paths has no significant relationship with the formation and break-up of ties within the network.

6 Conclusion

Literature suggests that knowledge production and scientific research are increasing conducted in collaborative work between different authors and institutions. Moreover, collaboration becomes increasingly more international, particularly in the pharmaceutical industry. In this study we analyzed international pharmaceutical research collaboration networks on the country level in different therapeutic areas. Our empirical analysis is

	Estimate	$\Pr(\leq b)$	$\Pr(\geq b)$	$\Pr(\geq \lvert b \rvert)$
Dependent Variable: ΔCollaborations				
DiffDegreeCentral	5.4525	0.9709	0.0291	0.0353
PrevExp	0.4189	1	0	0
SmIncomeGr	-2.3720	0.0011	0.9989	0.0196
PointConnectivity	0.1500	0.9987	0.0013	0.0014
GeoCount	-0.0480	0.1005	0.8995	0.2052
intercept	-0.3009	0.3111	0.6889	0.3154
Residual Standard Error	18.78			
Adjusted R^2	0.5827			
F-statistic (p-value)	1675	0		

Nullhypothesis: QAP with 10,000 permutations

Table 3: Network Regression

based on an unique dataset of journal publications related to pharmaceutical research. By means of social network analysis we find that the international research networks expand over time in almost all therapeutic areas. More precisely, the number of countries involved and their connectivity increases in most therapeutic areas. Visual inspection of the networks reveals that high income OECD countries are located in the core of all networks. This pattern remains rather stable over time and only few non-OECD countries manage to become part of the center of international pharmaceutical research.

In order to assess which mechanisms suggested by the literature, namely preferential attachment, homophily, or multi-connectivity, drive the endogenous network dynamics, we employ multiple regression analysis for dyadic data. The QAP procedure is used to evaluate the significance of the coefficients. Our preliminary regression results reveal a positive association between preferential attachment proxied by previous experience and differences in countries' degree centrality and the change in the number of collaborations. Homophily in terms of income groups is negatively related to the change in international collaborations. Multi-connectivity in terms of different countries connecting two actors is positively related, whereas the number of shortest path has no significant relation to the change in the number of collaborations.

A Appendix

A.1 List of Therapeutic Areas and Description of Variables

Therapeutic Area	Therapeutic Area ID
Cancer	1
Cardiovascular	2
Central Nervous System	3
Dermatology	4
Eye and Ear	6
Gastrointestinal	7
Genitourinary	8
Hematological	9
HIV Infections	10
Hormonal Systems	11
Immune System	12
Infectious Diseases	13
Musculoskeletal	15
Pain	16
Respiratory	17

Table 4: List of Therapeutic Areas

Dependent Variable		
ΔCollaborations		change in the number of collaborations among countries from period 2 to 3
Independent Variables		
DiffDegreeCentral	Preferential Attachment	difference in countries degree centrality lagged by one period
PrevExp	Preferential Attachment	cumulated number of collaborations among two countries in period 1 and 2
SmIncomeGr	Homophily	dummy indicating whether 2 countries belong to the same income group
PointConnectivity	Multiconnectivity	number of other countries that have to be removed in order to disconnect two actors lagged by one period
GeodesicCount	Multiconnectivity	number of shortest paths between two countries lagged by one period

Table 5: Overview of Variables

A.2 Network Correlations

	DiffDegreeCentral	PrevExp	SmIncomeGr	PointConnectivity	GeoCount
DiffDegreeCentral	1				
PrevExp	0.0944	1			
SmIncomeGr	0.1154	0.1405	1		
PointConnectivity	0.2956	0.2791	0.3300	1	
GeoCount	0.2369	-0.0123	0.1434	0.3442	1

Table 6: Network Correlations

References

Adams, J. D., Black, G. C., Clemmons, J. R., and Stephan, P. E. (2005). Scientific Teams and Institutional Collaborations: Evidence from U.S. Universities, 1981-1999. *Research Policy*, 34(3):259 – 285. University-based Technology Initiatives.

Ahuja, G. (2000). Collaboration Networks, Structural Holes, and Innovation: A Longitudinal Study. *Administrative Science Quarterly*, 45(3):425–455.

Arora, A. and Gambardella, A. (1994). The Changing Technology of Technological Change: General and Abstract Knowledge and the Division of Innovative Labour. *Research Policy*, 23(5):523 – 532.

Barabási, A.-L. (2003). Emergence of Scaling in Complex Networks. In Bornholdt, S. and Schuster, H. G., editors, *Handbook of Graphs and Networks: From the Genome to the Internet*, pages 69–84. Wiley-VCH, Weinheim.

Barabási, A.-L. and Albert, R. (1999). Emergence of Scaling in Random Networks. *Science*, 286(5439):509–512.

Boschma, R. and Frenken, K. (2010). The Spatial Evolution of Innovation Networks: A Proximity Perspective. In Boschma, R. and Martin, R., editors, *Handbook of Evolutionary Economic Geography*, chapter 5, pages 120–138. Edward Elgar Publishing, Cheltenham et al.

Boschma, R. A. and ter Wal, A. L. J. (2007). Knowledge Networks and Innovative Performance in an Industrial District: The Case of a Footwear District in the South of Italy. *Industry & Innovation*, 14(2):177 – 199.

Bozeman, B. and Corley, E. (2004). Scientists' Collaboration Strategies: Implications for Scientific and Technical Human Capital. *Research Policy*, 33(4):599–616.

Breschi, S. and Catalini, C. (2010). Tracing the Links Between Science and Technology: An Exploratory Analysis of Scientists' and Inventors' Networks. *Research Policy*, 39(1):14 – 26.

Butts, C. T. and Carley, K. M. (2001). Multivariate Methods for Inter-Structural Analysis. *CASOS Working Paper, Carnegie Mellon University.*

Calero, C., van Leeuwen, T. N., and Tijssen, R. J. W. (2007). Research Cooperation Within the Bio-pharmaceutical Industry: Network Analyses of Co-publications Within and Between Firms. *Scientometrics*, 71(1):8799.

Cantner, U. and Graf, H. (2006). The Network of Innovators in Jena: An Application of Social Network Analysis. *Research Policy*, 35(4):463 – 480.

Cooke, P. (2006). Global Bioregional Networks: A New Economic Geography of Bioscientific Knowledge. *European Planning Studies*, 14(9):1265 –1285.

de Nooy, W., Mrvar, A., and Batagelij, V. (2005). *Exploratory Social Network Analysis with Pajek*. Cambridge University Press, Cambridge.

de Solla Price, D. (1963). *Little Science, Big Science*. Columbia University Press, New York et al.

Dekker, D., Krackhardt, D., and Snijders, T. (2003). Multicollinearity Robust QAP for Multiple Regression. *CASOS Working Paper, Carnegie Mellon University*.

Dekker, D., Krackhardt, D., and Snijders, T. (2007). Sensitivity of MRQAP Tests to Collinearity and Autocorrelation Conditions. *Psychometrika*, 72:563–581. 10.1007/s11336-007-9016-1.

Freeman, L. C. (1977). A Set of Measures of Centrality Based on Betweenness. *Sociometry*, 40(1):35–41.

Freeman, L. C. (1979). Centrality in Social Networks Conceptual Clarification. *Social Networks*, 1(3):215 – 239.

Freeman, L. C. (1996). Some Antecedents of Social Network Analysis. *Connections*, 19(1):39–42.

Frenken, K. (2002). A New Indicator of European Integration and an Application to Collaboration in Scientific Research. *Economic Systems Research*, 14(4):345–361.

Fruchterman, T. M. and Reingold, E. M. (1991). Graph Drawing by Force-Directed Placement. *Software: Practice and Experience*, 21(11):11291164.

Gambardella, A. (1995). *Science and Innovation: The US Pharmaceutical Industry During the 1980s*. Cambridge University Press, Cambridge et al., 1 edition.

Gassmann, O. and von Zedtwitz, M. (1999). New Concepts and Trends in International R&D Organization. *Research Policy*, 28(2-3):231 – 250.

Gay, B. and Dousset, B. (2005). Innovation and Network Structural Dynamics: Study of the Alliance Network of a Major Sector of the Biotechnology Industry. *Research Policy*, Vol. 34(10):1457–1475.

Glückler, J. (2010). The Evolution of a Strategic Alliance Network: Exploring the Case of Stock Photography. In *Handbook of Evolutionary Economic Geography*. Edward Elgar Publishing Ltd, Cheltenham et al.

Granstrand, O. (1999). Internationalization of Corporate R&D: A Study of Japanese and Swedish Corporations. *Research Policy*, 28(2-3):275 – 302.

Guellec, D. and van Pottelsberghe de la Potterie, B. (2001). The Internationalisation of Technology Analysed with Patent Data. *Research Policy*, 30(8):1253 – 1266.

Hagedoorn, J. (2002). Inter-firm R&D Partnerships: An Overview of Major Trends and Patterns Since 1960. *Research Policy*, 31(4):477 – 492.

Hoekman, J., Frenken, K., and van Oort, F. (2009). The Geography of Collaborative Knowledge Production in Europe. *The Annals of Regional Science*, 43(3):721–738.

Hubert, L. and Schultz, S. (1976). Quadratic Assignment as a General Data Analysis Strategy. *British Journal of Mathematical and Statistical Psychology*, 29:190–241.

Hubert, L. J. (1987). *Assignment Methods in Combinatorial Data Analysis*. Dekker, New York et al.

Jackson, D. and Somers, K. (1989). Are Probability Estimates from the Permutation Model of Mantel's Test Stable? *Canadian Journal of Zoology*, 67(3):766–769.

Jeong, H., Néda, Z., and Barabási, A.-L. (2001). Measuring Preferential Attachment in Evolving Networks. *EPL (Europhysics Letters)*, 61(4):567–572.

Katz, J. S. and Martin, B. R. (1997). What is Research Collaboration? *Research Policy*, 26(1):1 – 18.

Kenis, P. and Knoke, D. (2002). How Organizational Field Networks Shape Interorganizational Tie-Formation Rates. *The Academy of Management Review*, 27(2):275–293.

Kim, J. W. and Higgins, M. C. (2007). Where do Alliances Come From?: The Effects of Upper Echelons on Alliance Formation. *Research Policy*, 36(4):499 – 514. Biotechnology: Its origins, organization, and outputs.

Knoke, D. and Burt, R. (1983). Prominence. In Burt, R. and Minor, M., editors, *Applied Network Analysis: A Methodological Introduction*, pages 195–222. Sage, Beverly Hills et al.

Kogut, B., Shan, W., and Walker, G. (1992). The Make or Cooperate Decision in the Context of an Industry Network. In Nohria, N. and Eccles, R., editors, *Networks and Organizations: Structure, Form and Action*, pages 348–365. Harvard Business School Press, Boston.

Krackhardt, D. (1987). QAP Partialling as a Test of Spuriousness. *Social Networks*, 9(2):171 – 186.

Krackhardt, D. (1988). Predicting with Networks: Nonparametric Multiple Regression Analysis of Dyadic Data. *Social Networks*, 10(4):359 – 381.

Laudel, G. (2002). What do we Measure by Co-authorships? *Research Evaluation*, 11(1):3–15.

Lundberg, J., Tomson, G., Lundkvist, I., Skar, J., and Brommels, M. (2006). Collaboration Uncovered: Exploring the Adequacy of Measuring University-Industry Collaboration Through Co-Authorship and Funding. *Scientometrics*, 69(3):575–589.

Mantel, N. (1967). The Detection of Disease Clustering and a Generalized Regression Approach. *Cancer Research*, 27(2):209–220.

Mattsson, P., Laget, P., Nilsson, A., and Sundberg, C.-J. (2008). Intra-EU vs. extra-EU scientific co-publication patterns in EU. *Scientometrics*, 75:555–574. 10.1007/s11192-007-1793-x.

McKelvey, M., Orsenigo, L., and Pammolli, F. (2004). Pharmaceuticals Analyzed Through the Lens of a Sectoral Innovation System. In Malerba, F., editor, *Sectoral Systems of Innovation: Concepts, Issues and Analyses of Six Major Sectors in Europe*. Cambridge University Press.

McPherson, J. M. and Smith-Lovin, L. (1987). Homophily in Voluntary Organizations: Status Distance and the Composition of Face-to-Face Groups. *American Sociological Review*, 52(3):370–379.

McPherson, M., Smith-Lovin, L., and Cook, J. M. (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, 27(1):415–444.

Mina, A., Ramlogan, R., Tampubolon, G., and Metcalfe, J. (2007). Mapping Evolutionary Trajectories: Applications to the Growth and Transformation of Medical Knowledge. *Research Policy*, 36(5):789 – 806.

Nagpaul, P. (2003). Exploring a Pseudo-regression Model of Transnational Cooperation in Science. *Scientometrics*, 56(3):403–416.

Newman, M. (2001). Clustering and Preferential Attachment in Growing Networks. *Physical Review E*, 64:1–4.

Okubo, Y. and Sjöberg, C. (2000). The Changing Pattern of Industrial Scientific Research Collaboration in Sweden. *Research Policy*, 29(1):81 – 98.

Orsenigo, L. (1989). *The Emergence of Biotechnology : Institutions and Markets in Industrial Innovation*. Pinter, London.

Orsenigo, L., Pammolli, F., Riccaboni, M., Bonaccorsi, A., and Turchetti, G. (1998). The Evolution of Knowledge and the Dynamics of an Industry Network. *Journal of Management and Governance*, 1(2):147–175.

Owen-Smith, J., Riccaboni, M., Pammolli, F., and Powell, W. W. (2002). A Comparison of U. S. and European University-Industry Relations in the Life Sciences. *Management Science*, 48(1):24–43.

Pisano, G. P. (1991). The governance of innovation: Vertical integration and collaborative arrangements in the biotechnology industry. *Research Policy*, 20(3):237 – 249.

Powell, W. and Brantley, P. (1992). Competitive Cooperation in Biotechnology: Learning Through Networks? In Eccles, R. and Nohria, N., editors, *Networks and Organizations*, pages 366–394. Harvard Business School Press, Boston.

Powell, W. and Grodal, S. (2005). Networks of Innovators. In Fagerberg, J., Mowery, D., and Nelson, R., editors, *The Qxford Handbook of Innovation*, pages 56–85. Oxford University Press, Oxford et al.

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

Powell, W., Koput, K., Smith-Doerr, L., and Owen-Smith, J. (1999). Network Position and Firm Performance: Organizational Returns to Collaboration in the Biotechnology Industry. In Andrews, S. and Knoke, D., editors, *Research in the Sociology of Organizations*, pages 129–159. JAI Press, Greenwich.

Powell, W., White, D., Koput, K., and Owen-Smith, J. (2005). Network Dynamics and Field Evolution: The Growth of Interorganizational Collaboration in the Life Sciences. *American Journal of Sociology*, Vol. 110(4):1132–1205.

Rogers, E. M. (1995). *Diffusion of Innovations*. Free Press, New York et al.

Roijakkers, N. and Hagedoorn, J. (2006). Inter-firm R&D Partnering in Pharmaceutical Biotechnology Since 1975: Trends, Patterns, and Networks. *Research Policy*, 35(3):431–446.

Ruef, M., Aldrich, H. E., and Carter, N. M. (2003). The Structure of Founding Teams: Homophily, Strong Ties, and Isolation among U.S. Entrepreneurs. *American Sociological Review*, 68(2):195–222.

Saxenian, A. (1991). The Origins and Dynamics of Production Networks in Silicon Valley. *Research Policy*, 20(5):423 – 437.

Schumpeter, J. A. (1912). *Theorie der wirtschaftlichen Entwicklung*. Duncker & Humblot, Leipzig.

Ter Wal, A. L. and Boschma, R. A. (2009). Applying Social Network Analysis in Economic Geography: Framing some Key Analytic Issues. *The Annals of Regional Science*, 43(3):739–756.

Tijssen, R. J. (2009). Measuring the Corporate Web of Science: Research and Partnership Networks Within the European Pharmaceutical Industry. In Malerba, F. and Vonortas, N. S., editors, *Innovation Networks in Industries*, pages 81–104. Edward Elgar Publishing, Cheltenham et al.

Verspagen, B. (2007). Mapping Technological Trajectories as Patent Citation Networks: A Study on the History of Fuel Cell Research. *Advances in Complex Systems*, 10(1):93–115.

Wagner, C. S. and Leydesdorff, L. (2005a). Mapping the Network of Global Science: Comparing International Co-authorships from 1990 to 2000. *International Journal of Technology and Globalisation*, 1(2):185–208.

Wagner, C. S. and Leydesdorff, L. (2005b). Network Structure, Self-organization, and the Growth of International Collaboration in Science. *Research Policy*, 34(10):1608 – 1618.

Wagner-Döbler, R. (2001). Continuity and Discontinuity of Collaboration Behaviour since 1800 from a Bibliometric Point of View. *Scientometrics*, 52(3):503–517.

Wasserman, S. and Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge University Press, Cambridge et al.

Watts, D. J. and Strogatz, S. H. (1998). Collective Dynamics of 'Small-world' Networks. *Nature*, 393(4):440–442.

Witt, U. (2008). Evolutionary Economics. In Durlauf, S. N. and Blume, L. E., editors, *The New Palgrave Dictionary of Economics*. Palgrave Macmillan, Basingstoke et al., 2nd edition.

Wuchty, S., Jones, B. F., and Uzzi, B. (2007). The Increasing Dominance of Teams in Production of Knowledge. *Science*, 316(5827):1036–1039.

Zucker, L. (1986). Production of Trust: Institutional Sources of Economic Structure, 1840-1920. *Research in Organizational Behavior*, 8:53–111.